January 2006

# HOSPITAL QUALITY DATA

## CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data

**G A O**

Accountability ★ Integrity ★ Reliability

# HOSPITAL QUALITY DATA

## CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data

## Why GAO Did This Study

The Medicare Modernization Act of 2003 directed that hospitals lose 0.4 percent of their Medicare payment update if they do not submit clinical data for both Medicare and non-Medicare patients needed to calculate hospital performance on 10 quality measures. The Centers for Medicare & Medicaid Services (CMS) instituted the Annual Payment Update (APU) program to collect these data from hospitals and report their rates on the measures on its Hospital Compare Web site.

For hospital quality data to be useful to patients and other users, they need to be reliable, that is, accurate and complete. GAO was asked to (1) describe the processes CMS uses to ensure the accuracy and completeness of data submitted for the APU program, (2) analyze the results of CMS's audit of the accuracy of data from the program's first two calendar quarters, and (3) describe processes used by seven other organizations that assess the accuracy and completeness of clinical performance data.

## What GAO Recommends

GAO recommends that CMS take steps to improve its processes for ensuring the accuracy and completeness of hospital quality data. In commenting on a draft of this report, CMS agreed to implement steps to improve the quality and completeness of the data.

www.gao.gov/cgi-bin/getrpt?GAO-06-54.

To view the full product, including the scope and methodology, click on the link above. For more information, contact Cynthia A. Bascetta, (202) 512-7101 or BascettaC@gao.gov.

## What GAO Found

CMS has contracted with an independent medical auditing firm to assess the accuracy of the APU program data submitted by hospitals, but has no ongoing process in place to assess the completeness of those data. CMS's independent audit checks accuracy by comparing the quality data submitted by hospitals from the medical records for a sample of five patients per calendar quarter for each hospital to the quality data that the contractor has reabstracted from the same records. The data are deemed to be accurate if there is 80 percent or greater agreement between these two sets of results. CMS has established no ongoing process to check data completeness. For the payment updates for fiscal years 2005 and 2006, CMS compared the number of cases submitted by a hospital to the number of Medicare claims that hospital submitted. However, these analyses did not address non-Medicare patient records, and the approach that CMS took in these analyses was not capable of detecting incomplete data for all hospitals.

Although GAO found a high overall baseline level of accuracy when it examined CMS's assessment of the data submitted for the first two quarters of the APU program, the results are statistically uncertain for up to one-third of hospitals, and a baseline level of data completeness cannot be determined. The median accuracy score of 90 to 94 percent—depending on the calendar quarter and measures used—was well above the 80 percent accuracy threshold set by CMS, and about 90 percent of hospitals met or exceeded that threshold for both the first and the second calendar quarters of 2004. However, for approximately one-fourth to one-third of all the hospitals that CMS assessed for accuracy, the statistical margin of error for their accuracy score included both passing and failing accuracy levels. Consequently, for these hospitals, the small number of cases that CMS examined was not sufficient to establish with statistical certainty whether they met the accuracy threshold set by CMS. With respect to completeness of data, CMS did not assess the extent to which all hospitals submitted data on all eligible patients, or a representative sample thereof, for the two baseline quarters. As a result, there were no data from which to derive an assessment of the baseline level of completeness of the quality data that hospitals submitted for the APU program.

Other reporting systems that collect clinical performance data have adopted a range of activities to ensure data accuracy and completeness, which include some methods employed by all, such as checking the data electronically to identify missing data. Officials from some of the other reporting systems and an expert in the field stressed the importance of including an independent audit in the methods used by organizations to check data accuracy and completeness. Most of the other reporting systems incorporate three methods into their process that CMS does not use in its independent audit. Specifically, most include an on-site visit in their independent audit, focus their audits on a selected number of facilities, and review a minimum of 50 patient medical records during the audit.

# Contents

**Figures**

---

---

**Abbreviations**

| | |
|---|---|
| ACC | American College of Cardiology |
| AMI | acute myocardial infarction |
| APU program | Annual Payment Update program |
| CABG | coronary artery bypass grafting |
| CAP | community-acquired pneumonia |
| CDAC | Clinical Data Abstraction Center |
| CMS | Centers for Medicare & Medicaid Services |
| DAVE | Data Assessment and Verification Project |
| HF | heart failure |
| HQA | Hospital Quality Alliance |
| IFMC | Iowa Foundation for Medical Care |
| JCAHO | Joint Commission on Accreditation of Healthcare Organizations |
| MDS | Minimum Data Set |
| MEDPAR | Medicare Provider Analysis and Review |
| MMA | Medicare Prescription Drug, Improvement, and Modernization Act |
| MSA | metropolitan statistical area |
| NCQA | National Committee for Quality Assurance |
| PCI | percutaneous coronary intervention |
| PTCA | percutaneous transluminal coronary angioplasty |
| QIO | quality improvement organization |
| SPARCS | Statewide Planning and Research Cooperative System |
| SSA | Social Security Administration |
| STS | Society of Thoracic Surgeons |

![GAO logo](Accountability * Integrity * Reliability)

**G A O**

**United States Government Accountability Office**
**Washington, DC 20548**

January 31, 2006

The Honorable Charles E. Grassley
Chairman
The Honorable Max Baucus
Ranking Minority Member
Committee on Finance
United States Senate

The Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003 created a financial incentive for hospitals to submit data to provide information about their quality of care that could be publicly reported.[1] Under Section 501(b) of MMA, acute care hospitals shall submit the clinical data from the medical records of all Medicare and non-Medicare patients needed to calculate hospitals' performance on 10 quality measures. If a hospital chooses not to submit the data, it will lose 0.4 percent of its annual payment update from Medicare for a subsequent fiscal year.[2] The Centers for Medicare & Medicaid Services (CMS) established the Annual Payment Update program (APU program)[3] to implement this provision of MMA. Participating hospitals submit quality data that are used to calculate a hospital's performance on the measures quarterly,[4] according to a schedule defined by CMS. MMA affects hospital annual payment updates for fiscal year 2005 through fiscal year 2007.[5] For fiscal year 2005, the first year of the program, CMS based its annual

---

[1]Pub. L. No. 108-173, § 501(b), 117 Stat. 2066, 2289-90 (amending section 1886(b)(3)(B) of the Social Security Act, to be codified at 42 U.S.C. § 1395ww(b)(3)(B)).

[2]The reduction in the annual payment update applies to hospitals paid under Medicare's inpatient prospective payment system. Critical access, children's, rehabilitation, psychiatric, and long-term-care hospitals may elect to submit data for any of the measures, but they are not subject to a reduction in their payment if they choose not to submit data.

[3]Throughout this report, we refer to CMS's Reporting Hospital Quality Data for the Annual Payment Update program as the "APU program".

[4]Throughout this report, we refer to the clinical data submitted by hospitals that are used to calculate their performance on the measures as "quality data".

[5]Senate Bill 1932 would extend the APU program indefinitely. It would also increase the penalty for not submitting data to 2 percent and provide for the Secretary to establish additional measures, beyond the original 10, for payment purposes.

payment update on quality data submitted by hospitals for patients discharged between January 1, 2004, and March 31, 2004.

Under MMA, the 10 quality measures for which hospitals report data are those established by the Secretary of Health and Human Services as of November 1, 2003. The measures cover three conditions: heart attack, heart failure, and pneumonia. Over 3 million patients were admitted to acute care hospitals in 2002 with these three conditions, representing approximately 10 percent of total acute care hospital admissions. For patients over 65, acute care hospital admissions for the three conditions represented approximately 16 percent of total admissions.

The collection of quality data on the 10 measures is part of a larger initiative to provide useful and valid information about hospital quality to the public.[6] In April 2005, CMS launched a Web site called "Hospital Compare" to convey information on these and other hospital quality measures to consumers. Additional measures are being introduced by CMS,[7] and it is expected that public reporting of hospital quality measures will continue into the future. Hospitals may submit quality data on additional measures for the APU program, but CMS bases any reduction in the annual payment update on the 10 measures referenced in the MMA. In addition to this effort, other public and private organizations also administer reporting systems in which clinical data are collected and may be released to the public.

In order for publicly released information on the hospital quality measures to be useful to patients, payers, health professionals, health care organizations, regulators, and other users, the quality data used to calculate a hospital's performance on the measures need to be reliable, that is, both accurate and complete. If a hospital submits complete data, that is, data on all the cases that meet the specific inclusion criteria for eligible patients, but the data are not collected, or abstracted, from the patients' medical records accurately, the data will not be reliable. Similarly, if a hospital submits accurate data, but those data are

---

[6]According to the Secretary of Health and Human Services, the effort is also intended to provide hospitals with a sense of predictability about public reporting expectations, to standardize data and data collection mechanisms, and to foster hospital quality improvement, in addition to providing information on hospital quality to the public.

[7]For example, CMS plans to publicly report on the Hospital Compare Web site measures of patient perspectives on seven aspects of hospital care, with national implementation scheduled for 2006.

incomplete because the hospital leaves out eligible cases, the data will not be reliable. Data that are not reliable may present a risk to people making decisions based on the data, such as a patient choosing a hospital for treatment. The program's initial, or baseline, data could describe data reliability at the start of the program and provide a reference point for any subsequent assessments.

You asked us to provide information on the reliability of publicly reported information on hospital quality obtained through the APU program. In this report, we (1) describe the processes CMS uses to ensure that the quality data submitted by hospitals for the APU program are accurate and complete and any plans by CMS to modify its processes; (2) determine the baseline levels of accuracy and completeness for the data for patients discharged from January 2004 through June 2004, the first two quarters of data submitted by hospitals under the APU program; and (3) describe the processes used by seven other organizations that collect clinical performance data to assess the accuracy and completeness of quality data for selected reporting systems.

In addressing these objectives, we collected information through interviews, examination of documents, and data analysis. To describe CMS's processes for ensuring the accuracy and completeness of the quality data for the APU program, we interviewed program officials from CMS and its contractors,[8] hospital associations, quality improvement organizations (QIO), and hospital data vendors.[9] In addition, we examined both publicly available and internal documents from CMS and its contractors. To determine the baseline accuracy and completeness of data submitted for the APU program, we drew on available information collected by CMS. In particular, we analyzed the accuracy of the quality data based on the reabstraction of patient medical records performed by

---

[8]CMS's contractors for this program are the Iowa Foundation for Medical Care (IFMC) and DynKePRO, LLC. IFMC is the quality improvement organization (QIO) for the state of Iowa. (QIOs are independent organizations that work under contract to CMS to monitor quality of care for the Medicare program and help providers to improve their clinical practices.) Under a separate contract, IFMC operates the national database for hospital quality data known as the QIO clinical warehouse. DynKePRO, LLC, an independent medical auditing firm, operates CMS's Clinical Data Abstraction Center (CDAC), which assesses the accuracy of hospital data submissions.

[9]Some hospitals contract with data vendors to electronically process, analyze, and transmit patient information.

CMS's Clinical Data Abstraction Center (CDAC).[10] The reabstraction results available at the time we conducted our analyses pertained to hospital discharges that took place from January 1, 2004, through June 30, 2004.[11] We extracted additional information about hospitals from the Medicare Provider of Services database, including the number of Medicare-certified beds and urban or rural location. After examining the CDAC data and reviewing the procedures that CMS has put in place to conduct the reabstraction process, we determined that the data were sufficiently reliable to use in estimating the baseline level of accuracy characterizing the quality data submitted by hospitals for those two calendar quarters. Regarding data on completeness of the quality data, we interviewed CMS officials and contractors and examined related documents. To examine the methods used by other reporting systems[12] to assess data completeness and accuracy, we conducted structured interviews with officials from seven organizations,[13] including government agencies, that administer such systems. We focused on reporting systems that collect clinical rather than administrative data. We selected a mix of systems, in terms of public or private sponsorship, types of providers assessed, and medical conditions covered, to ensure variety. We also spoke with individual health professionals with expert knowledge in the field of hospital quality assessment.

Our analysis of the level of accuracy and completeness of the quality data is based on the procedures developed by CMS to validate the data

---

[10]Reabstraction is the re-collection of clinical data for the purpose of assessing the accuracy of hospital abstractions. In the APU program, CDAC compares data originally submitted by the hospitals to those it has reabstracted from the same medical records.

[11]These were the calendar quarters for which, at the time we conducted our analysis, hospitals had collected the data and CMS had completed its process for reabstracting and assessing the data. We analyzed data for all hospitals affected by section 501(b) of MMA, which were located in 49 states and the District of Columbia. Hospitals in Maryland and Puerto Rico were excluded because they are paid under different payment systems than other acute care hospitals.

[12]Throughout this report, we refer to this group of quality data reporting systems, each of which collects some type of clinical performance data from designated providers or health plans, as "other reporting systems".

[13]The seven organizations were the American College of Cardiology, the California Office of Statewide Health Planning and Development, CMS (the units responsible for monitoring nursing home care regarding the Data Assessment and Verification Project contract), the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), the National Committee for Quality Assurance, the New York State Department of Health, and the Society of Thoracic Surgeons.

submitted; we have not independently compared the data submitted by hospitals to the original patient clinical records. In addition, we did not assess the performance of hospitals with respect to the quality measures themselves (which show how often the hospitals provided a specified service or treatment when appropriate). We conducted our work from November 2004 through January 2006 in accordance with generally accepted government auditing standards. For more details on our scope and methodology, see appendix I.

## Results in Brief

CMS has processes for ensuring the accuracy of the quality data submitted by hospitals for the APU program, but has no ongoing process for assessing the completeness of those data. To check accuracy, one CMS contractor electronically checks the data as they are submitted to the clinical warehouse, and another operates CMS's CDAC that conducts an independent audit by sampling five patient record abstractions from all the quality data submitted by each hospital in a quarter. CDAC then compares the quality data originally collected by the hospital from the medical records for those five patients to the quality data it has reabstracted from the same medical records. The data are deemed to be accurate if there is 80 percent or greater agreement between these two sets of results. CMS did not require hospitals to meet the 80 percent threshold for the 10 APU measures to receive their full annual payment update for fiscal year 2005. However, for fiscal year 2006, CMS reduced the payment update by 0.4 percentage points for hospitals whose data on the APU measures do not meet the 80 percent threshold. To assess completeness, CMS has twice compared the number of cases submitted by each hospital for the APU program for a given period to the number of claims each hospital submitted to Medicare, once for the fiscal year 2005 update and once for the fiscal year 2006 update. However, these analyses did not address non-Medicare patient records, and the approach that CMS took in these analyses was not capable of detecting incomplete data for all hospitals. For example, to determine which hospitals could receive the full fiscal year 2006 update, CMS limited its analysis to hospitals that submitted no patient data at all to the clinical warehouse in a given quarter. CMS has not put in place an ongoing process for checking the completeness of the data that hospitals submit for the APU program that would provide accurate and consistent information for all patients and all hospitals. Nor has CMS required hospitals to certify that they submitted data for all eligible patients or a representative sample thereof.

We could determine a baseline level of accuracy for the quality data submitted by hospitals for the APU program but not a baseline level of

completeness. We found a high overall baseline level of accuracy when we examined CMS's assessment of the data from the first two calendar quarters of 2004. Overall, the median accuracy score exceeded 90 percent, which was well above the 80 percent accuracy threshold set by CMS, and about 90 percent of hospitals met or exceeded that threshold for both the first and the second calendar quarters of 2004. For most hospitals whose accuracy score was well above the threshold, the results based on the reabstraction of five cases were statistically certain. However, for approximately one-fourth to one-third of all the hospitals that CMS assessed for accuracy, the statistical margin of error for their accuracy score included both passing and failing accuracy levels. Consequently, for these hospitals, the five cases that CMS examined were not sufficient to establish with statistical certainty whether the hospital met the threshold level of data accuracy. Accuracy did not vary between rural and urban hospitals, and small hospitals provided data as accurate as those from larger hospitals. The completeness baseline could not be determined because CMS did not assess the extent to which all hospitals submitted data on all eligible patients, or a representative sample thereof, for the first two calendar quarters of 2004, and consequently there were no data from which to derive such an assessment.

Other reporting systems that collect clinical performance data have adopted various methods to ensure data accuracy and completeness. Some of these methods are used by all of these other reporting systems, such as checking the data electronically to identify missing data. Officials from some of the other systems and an expert in the field stressed the importance of including an independent audit in the methods used by organizations to check data accuracy and completeness. Most other reporting systems that conduct independent audits incorporate three methods as part of their process that CMS does not use in its independent audit. Specifically, most include an on-site visit, focus their audits on a selected number of facilities or reporting entities, and review a minimum of 50 patient medical records per reporting entity during the audit.

In order for CMS to ensure that the hospital quality data are accurate and complete, we recommend that the CMS Administrator, focusing on the subset of hospitals for which it is statistically uncertain if they met CMS's accuracy threshold in one or more previous quarters, increase the number of patient records reabstracted by CDAC. We further recommend that CMS require hospitals to certify that they took steps to ensure that they submitted data on all eligible patients, or a representative sample thereof, and that the agency assess the level of incomplete data submitted by hospitals for the APU program to determine the magnitude of

underreporting, if any, in order to refine how completeness assessments may be done in future reporting efforts. In commenting on a draft of this report, CMS agreed to implement steps to improve the quality and completeness of the data.

# Background

Medicare spends over $136 billion annually on inpatient hospital care for its beneficiaries. To help ensure the quality of the care it purchases through Medicare, CMS launched the Hospital Quality Initiative in 2003. This initiative aims to refine and standardize hospital data, data transmission, and performance measures as part of an effort to stimulate and support significant improvement in the quality of hospital care.

One component of this broader initiative is CMS's participation in the Hospital Quality Alliance (HQA), a public-private collaboration that seeks to make hospital performance information more accessible to the public, payers, and providers of care.[14] Before the enactment of MMA, HQA had organized a voluntary program for hospitals to submit data on quality of care measures intended for public reporting. For its part as a participant in HQA, CMS set up a central database to receive the data submitted by hospitals and initiated plans for a Web site to post information on hospital quality of care measures. Thus, CMS had a data collection infrastructure in place when MMA established the financial incentive for hospitals to submit quality data.

## Selection of Measures

The 10 measures chosen by the Secretary of Health and Human Services for the APU program are the original 10 measures that were adopted by HQA. HQA subsequently adopted additional measures that relate to the same three conditions—heart attacks, heart failure, and pneumonia—and others that relate to surgical infection prevention. (See table 1 for a listing

---

[14]HQA (formerly called the National Voluntary Hospital Reporting Initiative) was initiated by the American Hospital Association, the Federation of American Hospitals, and the Association of American Medical Colleges. It is supported by CMS, as well as the Joint Commission on Accreditation of Healthcare Organizations, National Quality Forum, American Medical Association, Consumer-Purchaser Disclosure Project, AARP, AFL-CIO, and Agency for Healthcare Research and Quality. Its aim is to provide a single standard quality measure set for hospitals to support public reporting and pay-for-performance efforts.

of the APU-measure set and the expanded-measure set.[15]) Hospitals participating in HQA were encouraged to submit data on the additional measures, but data submitted on the additional measures did not affect whether a hospital received its full payment update under the APU program. CMS and the QIOs have tested these measures for validity and reliability, and all measures have been endorsed by the National Quality Forum, which fosters agreement on national standards for measurement and public reporting of health care performance data.[16]

---

[15]Throughout this report, we refer to the 10 measures on which reductions in the annual payment update are based as the "APU-measure set" and to the combination of those 10 with the additional measures adopted by HQA as the "expanded-measure set". HQA added 7 measures for discharges beginning April 1, 2004, and another 5 measures for discharges beginning July 1, 2004, for a total of 22 measures on which hospitals may currently submit data. Thus, the expanded-measure set includes different numbers of measures for different quarters of data.

[16]The National Quality Forum is a voluntary standard-setting, consensus-building organization representing providers, consumers, purchasers, and researchers.

**Table 1: HQA Hospital Quality Measures**

| | Heart attack | Heart failure | Pneumonia | Surgical infection prevention |
|---|---|---|---|---|
| **APU-measure set** | | | | |
| *For discharges beginning January 1, 2004* | 1. Aspirin at arrival <br> 2. Aspirin prescribed at discharge <br> 3. ACE (angiotensin-converting enzyme) inhibitor for left ventricular systolic dysfunction <br> 4. Beta blocker at arrival <br> 5. Beta blocker prescribed at discharge | 6. Left ventricular function assessment <br> 7. ACE inhibitor for left ventricular systolic dysfunction | 8. Initial antibiotic received within 4 hours of hospital arrival <br> 9. Oxygenation assessment <br> 10. Pneumococcal vaccination status | (none) |
| **Expanded-measure set** | | | | |
| *For discharges beginning April 1, 2004* | 1-5 above plus <br> 11. Thrombolytic agent received within 30 minutes of hospital arrival <br> 12. PTCA (percutaneous transluminal coronary angioplasty) received within 90 minutes of hospital arrival <br> 13. Adult smoking cessation advice/counseling | 6-7 above plus <br> 14. Discharge instructions <br> 15. Adult smoking cessation advice/counseling | 8-10 above plus <br> 16. Blood culture performed before first antibiotic received in hospital <br> 17. Adult smoking cessation advice/counseling | (none) |
| *For discharges beginning July 1, 2004* | 1-5, 11-13 above | 6-7, 14-15 above | 8-10, 16-17 above plus <br> 18. Initial antibiotic selection for CAP (community-acquired pneumonia) in immunocompetent patient <br> 19. Influenza vaccination[a] | 20. Prophylactic antibiotic received within 1 hour prior to surgical incision <br> 21. Prophylactic antibiotic selection for surgical patients[a] <br> 22. Prophylactic antibiotics discontinued within 24 hours after surgery end |

Source: CMS, as of August 4, 2005.

Note: Measures are worded as CMS posted them on www.qnetexchange.org.

[a]Hospitals are collecting data for these measures, but public reporting of hospital performance on these measures has been postponed.

To minimize the data collection burden on hospitals by the APU program, CMS and the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) have worked to align their procedures and protocols for collecting and reporting the specific clinical information that is used to score hospitals on the measures. JCAHO-accredited hospitals—approximately 82 percent of hospitals that participate in Medicare—have since 2002 submitted data to JCAHO on the same measures as those in the APU-measure set as well as many of those in the expanded-measure set. Beginning with the first calendar quarter of data submitted by hospitals for the APU program, hospitals had the option of submitting the same data to CMS that many of them were already collecting for JCAHO. In November 2004, CMS and JCAHO jointly issued a manual laying out the aligned procedures and protocols for discharges beginning January 1, 2005.

## Collection, Submission, and Reporting of Quality Data

Hospitals use CMS's definition of the eligible patient population to identify the patients for whom they should collect and submit quality data for each measure. The definition is based on the primary diagnosis and, for the two cardiac conditions, the age of the patient.[17] Specifically, hospitals use diagnostic codes and demographic information from the patients' medical and administrative records to determine eligibility based on protocols established by CMS.

Once the eligible patients have been identified, hospitals extract from their patients' medical records the specific data items needed for the Iowa Foundation for Medical Care (IFMC) to calculate a hospital's performance, following detailed data abstraction guidelines developed by CMS. Hospitals may submit data for all eligible patients for a given condition, or if they have more than a specified number of eligible patients, they may draw a random sample according to a formula,[18] and submit data for those

---

[17]Patients under 18 years of age are excluded from the eligible patient population for the two cardiac conditions.

[18]Before hospitals can consider sampling, rather than submitting all of their eligible cases, the number of eligible cases must exceed a minimum sample size that ranges from 60 per quarter for pneumonia cases to 76 for heart failure cases and 78 for heart attack cases. Once hospitals reach that threshold for a given condition, they can submit a random sample of their cases as long as the minimum sample size is met and it includes at least 20 percent of their eligible cases, up to a maximum sample size requirement of 241 for pneumonia, 304 for heart failure, and 311 for heart attacks. For discharges that occurred prior to January 1, 2005, CMS applied a different formula to hospitals not accredited by JCAHO that called for a minimum sample size of 7 for each of the three conditions and a sampling rate of at least 20 percent until a maximum sample size requirement of 70 cases was reached.

patients only. These data are put into a standardized data format and submitted quarterly through a secure Internet connection to the QIO "clinical warehouse" administered by IFMC. IFMC accepts into the clinical warehouse only the data that meet the formatting and other specifications established by CMS[19] and that are submitted before the specified deadline for that quarter. About 80 percent of hospitals rely on data vendors—which typically are collecting the same data for JCAHO—to submit the data for them.

IFMC aggregates the information from the individual patient records to generate a rate for each hospital on each of the measures for which the hospital submitted relevant clinical data. These rates show how often a hospital provided the specific service or activity designated in the measures to patients for whom that service or activity was appropriate. Hospitals also collect information on each patient that identifies patients for whom the particular service or activity would not be called for, such as patients with a condition that would make prescribing aspirin or beta blockers medically inappropriate.

CMS posts on its Hospital Compare Web site each hospital's rates for all the APU and expanded measures for which it submitted data.[20] In November 2004, CMS first posted these rates, based on data from the first quarter of calendar year 2004. It subsequently posted new rates in March 2005, based on the first two quarters of calendar year 2004 data, and again in September and December 2005 with additional quarters of data. CMS continues to update these rates quarterly, using the four most recent quarters of data available. There can be up to a 14-month time lag between when patients are treated by the hospital and when the resulting rates are posted on the CMS Web site. (See fig. 1.)

---

[19]IFMC statistics show that a majority of hospitals ultimately succeed in gaining acceptance for all the cases they have submitted and that less than 10 percent of hospitals have had more than 5 percent of their cases rejected in a given quarter.

[20]For two measures, influenza vaccination and prophylactic antibiotic selection for surgical patients, CMS has postponed public reporting.

**Figure 1: Approximate Times for Collection, Submission, and Reporting of Hospital Quality Data**
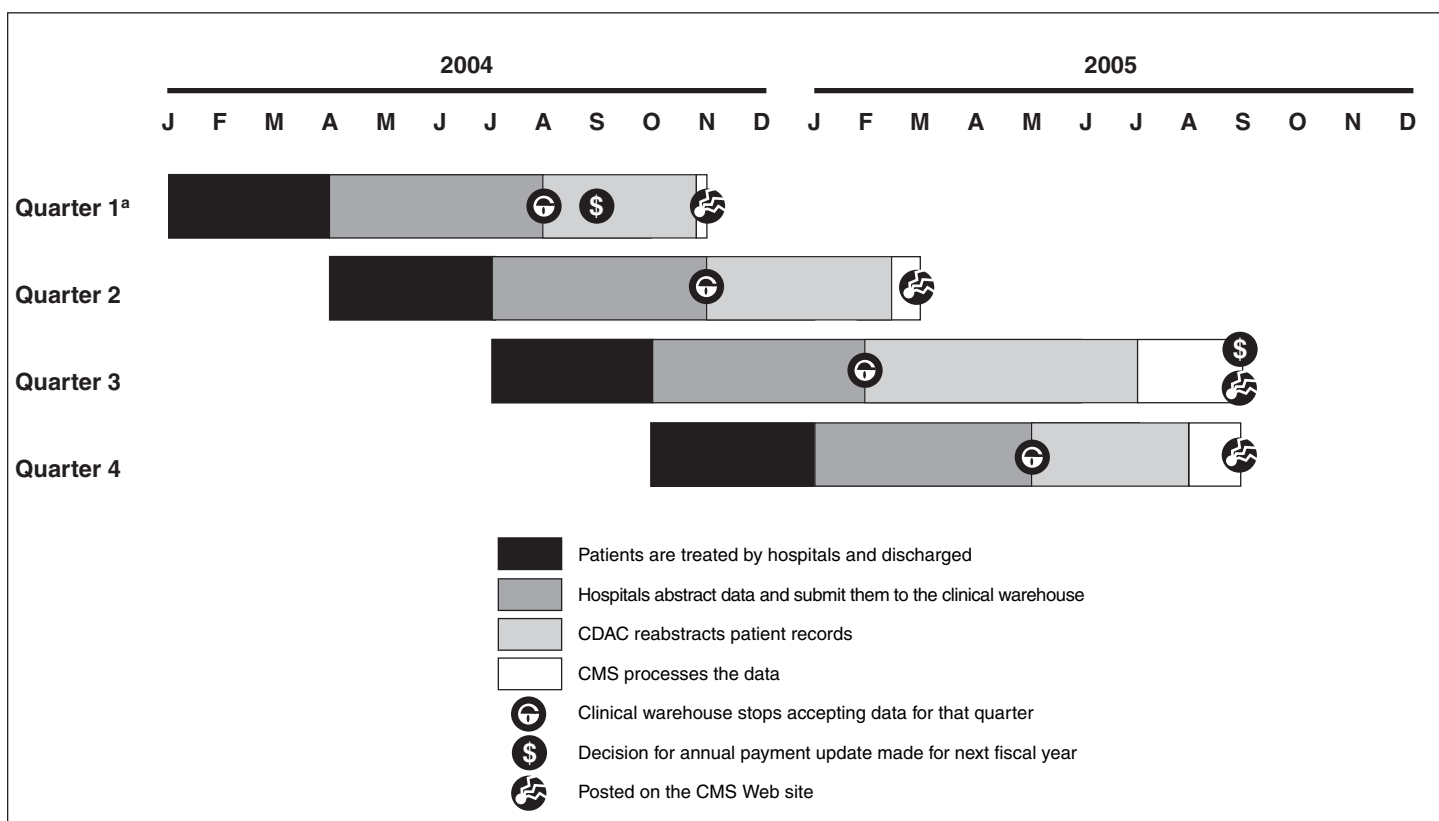


Source: GAO analysis based on CMS documents.

[a]CMS had to make its determination of hospital eligibility for the fiscal year 2005 annual payment update decision approximately 1 month after hospitals submitted their data for the first quarter.

| Implementation of the APU Program | In implementing the APU program, CMS uses the same policies and procedures for collecting and submitting quality data as are used for HQA. For the first annual payment update determined by the APU program, which applied to fiscal year 2005, hospitals were required to begin submitting data by July 1, 2004, for the patients discharged during the first calendar quarter of 2004 (January through March 2004). Data were received from 3,839 hospitals, over 98 percent of those affected by the MMA provision. These figures include 150 hospitals that certified to CMS that they had no eligible patients with the three conditions during the first calendar quarter of 2004. Hospitals that have no eligible patients are not penalized and receive the full annual payment update. For the second annual payment update determined by the APU program, which applied to fiscal year 2006, participating hospitals were required to continue to |
| --- | --- |

submit data in accordance with the quarterly deadlines set by CMS. Failure to meet the requirements of the program and qualify for the full annual payment update in one year does not affect a hospital's ability to participate in and qualify for the full update in the succeeding year.

CMS has assigned primary responsibility to the 53 QIOs to inform hospitals about the APU program's requirements and to provide technical assistance to hospitals in meeting those requirements. This includes assistance to hospitals in submitting their data to the clinical warehouse provided by IFMC.

## Other Reporting Systems

There are several organizations that administer reporting systems that collect clinical data, some of which also release their data to the public. Some of these organizations are in the public sector, such as state health departments, and some are in the private sector, such as accreditation bodies. Several of these systems have been in existence for a number of years, including one for as long as 16 years. Hospitals, health plans, nursing homes, and other external organizations submit data to these systems on a range of medical conditions, which for most of these systems includes at least one cardiac condition (e.g., percutaneous coronary intervention, coronary artery bypass grafting, heart attack, heart failure). Many of these systems make the results of the data they have collected available for public use. For example, one public organization has been collecting individual, patient-level data on cardiac surgeries from hospitals for the past 16 years and creates reports based on the data collected, which it subsequently posts on its Web site. Additionally, data collected by these reporting systems can also be used for quality improvement efforts and to track performance over time. (For more background information on other reporting systems, see app. II, table 3.)

## CMS Has Processes for Checking Data Accuracy but Has No Ongoing Process to Check Completeness

CMS has processes for ensuring the accuracy of the quality data submitted by hospitals for the APU program, but has no ongoing process to assess whether hospitals are submitting complete data. To check accuracy, IFMC, a CMS contractor, electronically checks the data as they are submitted to the clinical warehouse. In addition, CDAC independently audits the data submitted by hospitals. Specifically, it reabstracts the quality data from medical records for a sample of five patients per quarter for each hospital and compares its results to the quality data submitted by hospitals. The data are deemed to be accurate if there is 80 percent or greater agreement between these two sets of results, a standard that hospitals had to meet for the APU-measure set to qualify for their full annual payment update for

fiscal year 2006. To check completeness, CMS has twice compared the number of cases submitted by each hospital for the APU program for a given period to the number of claims the hospital submitted to Medicare, once for the fiscal year 2005 update and once for the fiscal year 2006 update. However, these analyses did not address non-Medicare patient records and the approach that CMS took in these analyses was not capable of detecting incomplete data for all hospitals. CMS has not put in place an ongoing process for checking the completeness of the data that hospitals submit for the APU program that would provide accurate and consistent information for all patients and all hospitals. Moreover, CMS has not required hospitals to certify that they submitted data for all eligible patients or a representative sample thereof.

## CMS Checks Data Accuracy Electronically and Through an Independent Audit

CMS employs two processes to check and ensure the accuracy of the quality data submitted by hospitals for the APU program. First, at the time that data are submitted to the clinical warehouse, IFMC, a CMS contractor, electronically checks the data for inconsistencies and missing values. The results are shared with hospitals. After the allotted time for review and correction of the submissions, no more data or corrections may be submitted by hospitals for that quarter. These checks are done whether the hospital submits its data directly to the warehouse or through a data vendor.

Second, CDAC conducts quarterly independent audits to verify that the data submitted by hospitals to the clinical warehouse accurately reflect the information in their patients' medical records.[21] From among all the patient records submitted to the clinical warehouse each quarter, CMS randomly selects for CDAC's reabstraction five patient records from each participating hospital.[22] CDAC sends a request for these patients' medical records to the hospitals, and they send photocopies of the records to CDAC for reabstraction. A CDAC abstractor reviews the medical record, determines if or when a specific action occurred—such as the time when a patient arrived at the hospital—and records that data field accordingly. Once the CDAC reabstraction is complete, the response previously entered

[21]DynKePRO, LLC, has operated CDAC since 1994. For 10 years it shared this function with a second firm, but in September 2004 DynKePRO negotiated a new contract with CMS that made it the sole CDAC contractor. In April 2005, DynKePRO became CSC York.

[22]To be included in the reabstraction process, hospitals must have submitted data on at least six patients across all three conditions in that quarter.

into that field by the hospital is compared to that entered by the CDAC abstractor, and CDAC notes whether the two responses match. If they do not match, a second CDAC abstractor reviews the medical record to make a final determination. The results of the CDAC reabstraction are sent to the clinical warehouse, where the individual data matches and mismatches are summed to produce an accuracy score for each hospital. The accuracy score represents the overall percentage of agreement between data submitted by the hospital and data reabstracted by CDAC across all five cases.[23] It is based on all the APU and expanded measures for which the hospital submitted data.[24] The score, along with information from CDAC on where the mismatches occurred and why, is shared with the hospital and the hospital's local QIO. CMS considers hospitals achieving an accuracy score of 80 percent or better to have provided accurate data. Hospitals with accuracy scores below 80 have the opportunity to appeal their reabstraction results.[25]

In applying these processes for the fiscal year 2005 annual payment update, CMS did not require hospitals to meet the 80 percent accuracy threshold for the 10 APU measures to qualify for the full update. Rather, to receive their full payment update, hospitals only had to pass the electronic data checking performed when they submitted their data to the clinical warehouse for the first calendar quarter of the APU program—for discharges that occurred from January 2004 through March 2004. Although

---

[23]The accuracy score is not based on all the data submitted by a hospital. Rather, CMS has identified a specific subset of the data elements that should be counted in computing the accuracy score. In general, CMS included in this subset the clinical data elements needed to calculate the hospital's rate for each of the measures and left out other administrative and demographic information about the patients. CMS estimates that five patient records usually contain about 100 data elements for calculation of the accuracy score, but the actual number of data elements depends on which conditions were involved and the number of measures for which a hospital submitted data.

[24]Although CMS computes accuracy scores based on data for all measures submitted to the clinical warehouse, it recognizes that the MMA provision affecting hospital payments applies only to data for the 10 measures specified for the APU program. See 69 *Fed. Reg.* 49080 (Aug. 11, 2004).

[25]CMS created an appeal process that allows a hospital to challenge the reabstraction results through its local QIO. For data from the first two calendar quarters of 2004, if the QIO agreed with the hospital's interpretation, the appeal was forwarded to CDAC for review and correction, if appropriate. CDAC's decision on the appeal was final. Beginning with data from the third calendar quarter of 2004, appealed cases no longer go back to CDAC. Instead, QIOs make the final decision to uphold either CDAC's or the hospital's interpretation. During this process, hospitals are not allowed to supplement the submitted patient medical records.

the accuracy scores were not considered for the payment update, CMS calculated an accuracy score for each quarter in which the hospital submitted at least six cases to the clinical warehouse. Each quarter the accuracy score was based on data for all the measures submitted by the hospital in that quarter and was derived from five randomly selected patient records. Along with the accuracy score, hospitals received information on where mismatches occurred and the reasons for the mismatches.

In contrast to the prior year, CMS applied the 80 percent threshold for accuracy as a requirement for hospitals to qualify for their full fiscal year 2006 annual payment update.[26] IFMC continued to check electronically all of the data as they were submitted for each quarter and calculated accuracy scores quarterly for each hospital. CMS decided to base its payment update decision on the accuracy score that hospitals obtained for the third calendar quarter of 2004—for discharges that occurred from July 2004 through September 2004.[27] This meant that the payment decision rested on the reabstraction results obtained from 5 randomly selected patient records. If a hospital met the 80 percent accuracy threshold based on all of the quality data it submitted, it received the full payment update. However, if a hospital failed to meet the 80 percent threshold, CMS recomputed the accuracy score using only the data elements required for the APU-measure set. For hospitals that failed again, CMS combined the CDAC reabstraction results from the third calendar quarter of 2004 with the CDAC results from the fourth calendar quarter of 2004 to produce an accuracy score derived from 10 patient medical records.[28] CMS then computed accuracy scores first for all the quality data submitted by the hospital and finally for the APU-measure set, if needed to reach the 80 percent threshold. As a result, even though CMS assessed hospital accuracy primarily on the basis of data that exceeded those required for the APU-measure set, hospitals were not denied the full annual payment

---

[26]70 *Fed. Reg.* 47420-47428 (Aug. 12, 2005).

[27]CMS decided not to use accuracy scores from the first two quarters of the APU program because those data were collected before the alignment of CMS and JCAHO data collection specifications had begun to come into effect. Given the time needed to conduct all the steps in the process (see fig. 1), CMS was left with the third calendar quarter of 2004 as the latest full quarter of data that could be used for determining the fiscal year 2006 update. The third calendar quarter also marked HQA's expansion to 22 measures.

[28]Hospitals had to submit their patient medical records to CDAC for the fourth calendar quarter 2004 reabstractions no later than August 1, 2005, to take advantage of this additional opportunity to pass the 80 percent threshold.

update except on the basis of the APU-measure set. A possibility does exist, however, that a hospital could have qualified for the full update based on its results for all the data it submitted, even if it would have failed using the APU-measure set. This could happen if the hospital submitted data that matched the CDAC abstractors' entries more consistently for the data entries used exclusively in computing the expanded measures, such as those relating to smoking cessation counseling, than for the data required by the APU-measure set.

In the future, CMS intends to base its decisions on hospital eligibility for full annual payment updates on accuracy assessments from more than one quarter. Although its concerns about potential alignment issues affecting data for the first two quarters of the APU program led the agency to rely primarily on data from the third calendar quarter for the fiscal year 2006 update, CMS stated that its goal was to use accuracy assessments from four consecutive quarters when it determines hospital eligibility for the fiscal year 2007 full annual payment update.

CMS uses the accuracy scores in making decisions on payment updates, but the scores do not affect the information posted on the Hospital Compare Web site. The Web site transmits to the public the rates on the APU and expanded measures that derive from the data that the hospitals submitted to the clinical warehouse. CMS does not post the accuracy scores generated from the CDAC reabstraction process on the Web site or indicate if the hospital rates are based on data that met CMS's 80 percent threshold for accuracy.[29]

## CMS Has No Ongoing Process to Ensure Completeness of Data Submitted for the APU Program

Although CMS has recognized the importance of obtaining quality data for the APU program on all eligible patients, or a representative sample if appropriate, it has not put in place an ongoing process to ensure that this occurs. For the fiscal year 2005 annual payment update, CMS checked that hospitals submitted data for at least a minimum number of patients by using Medicare claims data to estimate the number of "expected cases" that each hospital should have submitted to the clinical warehouse. To do this, it first calculated the average number of patients for each of the three conditions that each hospital had billed Medicare for over the previous

---

[29]The Hospital Compare Web site identifies instances where rates for a measure were based on fewer than 25 cases and where data were suppressed due to inaccuracies. However, the latter indication reflects situations where a hospital had problems with transmission of its data by a data vendor, not the outcome of the CDAC reabstractions.

eight calendar quarters (January 2002 through December 2003). Then, if the average number of Medicare claims for a condition was large enough to entitle the hospital to draw a sample instead of submitting data for all the eligible patients to the clinical warehouse, CMS reduced the number of "expected cases" based on the size of the sample.[30] CMS told each hospital what its expected numbers of heart attack, heart failure, and pneumonia patients were. If the actual number of patients for whom hospitals submitted data for the APU program was lower, the hospitals were instructed to send a letter to their local QIO, signed by the hospital's CEO or administrator, stating that the hospital had fewer discharged patients for that condition than CMS had estimated. If such a letter was filed, the hospital qualified for the full annual payment update. In the end, no hospital participating in the APU program was denied a full annual payment update for fiscal year 2005 for submitting data on an insufficient number of patients or any other reason.

For the fiscal year 2006 update decision, CMS took a different approach to using Medicare claims data to address the issue of completeness. CMS used Medicare claims data to check whether hospitals that billed Medicare for any cases with one of the three conditions submitted at least one case to the clinical warehouse. To do this, CMS compared each hospital's Medicare claims for the three conditions for the four calendar quarters of 2004 to the hospital's submissions to the clinical warehouse for those same quarters. CMS identified instances where hospitals had submitted one or more claims for payment to Medicare for any of the three conditions for a quarter when they had not submitted any cases with one of those conditions to the clinical warehouse. On this basis, CMS determined that 110 hospitals would not qualify for the full payment update for fiscal year 2006.

CMS conducted two additional analyses involving a comparison of the same Medicare claims data and quality data submissions to identify hospitals that may have submitted incomplete data for the APU program,

---

[30]Originally, CMS intended to apply JCAHO's sampling rules to JCAHO-accredited hospitals, and its own sampling rules to the other hospitals, in computing their "expected cases". JCAHO's sampling procedures called for submitting larger samples to the clinical warehouse than CMS's did. However, when CMS officials determined that they could not reliably identify every hospital that belonged in the JCAHO group, they decided to apply the CMS rules across the board to all hospitals. Therefore, for many JCAHO-accredited hospitals, the number of "expected cases" computed by CMS underestimated the number of Medicare cases for which these hospitals should have submitted data, because JCAHO-accredited hospitals were to submit cases according to the JCAHO sampling rules.

but these analyses did not affect hospital eligibility for the full fiscal year 2006 payment update. The additional analyses identified (1) a set of hospitals that may have submitted samples of their eligible cases to the clinical warehouse when, according to the applicable sampling rules, they should have submitted data on all their cases; and (2) another set of hospitals that failed to submit cases to the clinical warehouse for all of the three conditions for which they filed Medicare claims in that quarter. However, in contrast to the hospitals that did not qualify for their full payment update, the hospitals in the second set submitted to the clinical warehouse at least one case for one of the three conditions. A CMS official stated that the agency plans to educate the hospitals identified by these additional analyses on the data submission and sampling requirements for the APU program.

The analysis that CMS conducted using Medicare claims data for its fiscal year 2005 update decision and the three analyses it conducted in conjunction with its fiscal year 2006 update decision shared two limitations: none addressed the completeness of data submissions for non-Medicare patients, and none could detect incomplete data for all hospitals. Given that non-Medicare patients represent a substantial proportion of the patients treated for heart attacks, heart failure, and pneumonia,[31] any minimum number of "expected cases" based on Medicare claims inherently underestimates the total number of patients for which hospitals should have submitted quality data for the APU program. Moreover, the approaches taken in the analyses conducted for both fiscal year updates could not detect incomplete data for many hospitals. For example, in the fiscal year 2005 analysis, the difference between the number of cases expected under the CMS sampling rules and the higher number expected under the sampling rules that applied to JCAHO-accredited hospitals meant that JCAHO-accredited hospitals treating more patients than the minimum CMS sample of seven could have failed to submit data on most of the cases that exceeded the CMS minimum and still have met the number of expected cases set by CMS.[32] The analysis that CMS conducted to determine hospital eligibility for the full fiscal year 2006 update also could identify only certain hospitals that submitted incomplete data, in

---

[31]Non-Medicare patients account for about 40 to 50 percent of all patients hospitalized for heart attacks and pneumonia and 20 to 32 percent of those hospitalized for heart failure. For individual hospitals, these percentages could be higher or lower.

[32]See appendix I for more detailed information on the limitations that applied to CMS's effort to estimate a minimum number of expected cases for each hospital.

this case limited to hospitals that submitted no patient data at all to the clinical warehouse in a given quarter.

CMS officials acknowledged that the lack of information on non-Medicare patients and the imprecise adjustments that CMS made to take account of the varying sampling procedures that hospitals could have followed limited the conclusions that CMS could draw from its Medicare claims data analysis for the fiscal year 2005 update. Because of these limitations, CMS officials described their effort as a rough check for inconsistencies between data submitted by hospitals to the clinical warehouse and the cases that the hospitals had billed to Medicare.

CMS has not combined these limited efforts to monitor the completeness of hospital quality data submissions with efforts to clearly inform hospital officials of their obligation to submit complete data. For example, CMS has not explicitly listed submission of complete data as a requirement for participating in the APU program on the "Notice of Participation" that the hospital CEO or administrator must sign when hospitals enroll. The notice states requirements for participating hospitals—including that they must register with the QualityNet Exchange Web site[33] and that they must submit data for all measures specified in the APU-measure set by established deadlines. The notice indicates that the submitted data will undergo validation, a reference to the CDAC reabstraction process. However, the notice does not stipulate that hospitals must submit data for all eligible cases, or for a representative sample if appropriate.

We interviewed health professionals familiar with the APU program, several of whom raised concerns about data completeness. One expert in the area of outcomes research noted the potential for systematic underreporting by hospitals. He suggested that, as one approach to detect systematic underreporting, CMS could compare not only the number of patients for whom data were submitted and Medicare claims filed, but also the characteristics of patients for cases submitted to the APU program to the patient characteristics of comparable cases submitted to Medicare for payment. Another expert in the area of clinical quality improvement expressed his concern that the APU program did not verify the completeness of the data. He observed that hospitals have flexibility in determining which patients are included through their assignment of the

---

[33]The QualityNet Exchange Web site is the secure Internet connection used to transmit hospital quality data to the clinical warehouse.

**GAO-06-54 Reliability of Hospital Quality Data**

patient's primary diagnosis. A QIO official echoed this concern, noting the risk that hospitals could decide to not submit cases where patients had not received the services or activities assessed by the APU measures.

## Data Accuracy Baseline Was High Overall, but Statistically Uncertain for Many Hospitals, and Data Completeness Baseline Cannot Be Determined

We could determine a baseline level of accuracy for the quality data submitted for the APU program but not a baseline level of completeness. We found a high overall baseline level of accuracy when we examined CMS's assessment of the data submitted by hospitals for the first two calendar quarters of 2004. The median accuracy score exceeded 90 percent, which was well above the 80 percent accuracy threshold set by CMS, and about 90 percent of hospitals met or exceeded that threshold for both the first and the second calendar quarters of 2004. For most hospitals whose accuracy scores were well above the threshold, the results were statistically certain. However, for approximately one-fourth to one-third of all the hospitals that CMS assessed for accuracy, the statistical margin of error for their accuracy score included both passing and failing accuracy levels. Consequently, for these hospitals, the small number of cases that CMS examined was not sufficient to establish with statistical certainty whether the hospital met the threshold level of data accuracy. Accuracy did not vary between rural and urban hospitals, and small hospitals provided data as accurate as those from larger hospitals. The completeness baseline could not be determined because CMS did not assess the extent to which all hospitals submitted data on all eligible patients, or a representative sample thereof, for the first two calendar quarters of 2004, and consequently there were no data from which to derive such an assessment.

## Baseline Level of Data Accuracy Was High Overall, and Large Majority of Hospitals Met Accuracy Threshold

Overall, the baseline level of data accuracy for the first two quarters of the APU program was high. The median accuracy score achieved by hospitals ranged between 90 and 94 percent, with slightly higher values in the second quarter and for the APU-measure set. (See fig. 2.) In addition, with at least half the hospitals receiving accuracy scores above 90, relatively few failed to reach the 80 percent threshold set by CMS.

**Figure 2: Baseline Hospital Accuracy Scores at Selected Percentiles, by Measure Set and Quarter**
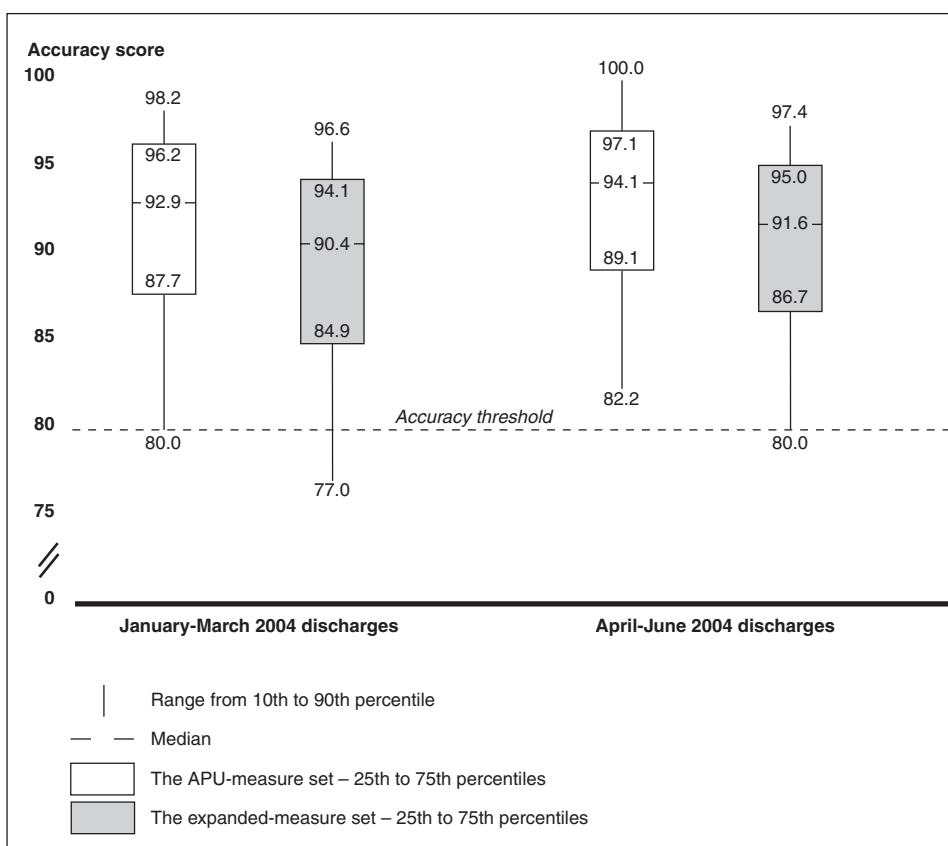


Source: GAO analysis of CMS data.

Note: Figure reflects accuracy scores for hospitals covered by the APU program. Hospitals that submitted fewer than six cases to the clinical warehouse in a quarter did not undergo CDAC reabstraction and therefore did not receive an accuracy score for that quarter. Calculation of accuracy scores for the expanded-measure set was based on all the measures for which a hospital submitted data, which could range from the APU measures alone to a maximum of 17—the APU measures plus as many as 7 additional measures.

In both quarters, 90 to 92 percent of hospitals obtained accuracy scores meeting the threshold using the APU-measure set, and 87 to 90 percent met the threshold using the expanded-measure set (see table 2).[34] The 8 to 13 percent of hospitals that did not meet the accuracy threshold represented approximately 300 to 500 hospitals across the country.

**Table 2: Percentage and Number of Hospitals Whose Baseline Accuracy Score Met or Fell Below the 80 Percent Threshold, by Measure Set and Quarter**

| | APU-measure set | | | | Expanded-measure set | | | |
|---|---|---|---|---|---|---|---|---|
| | January-March 2004 discharges | | April-June 2004 discharges | | January-March 2004 discharges | | April-June 2004 discharges | |
| | Percentage | Number | Percentage | Number | Percentage | Number | Percentage | Number |
| Hospitals whose accuracy score met 80 percent threshold | 90.2 | 3,290 | 91.8 | 3,282 | 86.8 | 3,165 | 90.0 | 3,217 |
| Hospitals whose accuracy score fell below 80 percent threshold | 9.8 | 359 | 8.2 | 292 | 13.2 | 483 | 10.0 | 357 |
| **Total** | **100** | **3,649** | **100** | **3,574** | **100** | **3,648** | **100** | **3,574** |

Source: GAO analysis of CMS data.

Note: Calculation of accuracy scores for the expanded-measure set was based on all the measures for which a hospital submitted data, which could range from the APU measures alone to a maximum of 17—the APU measures plus as many as 7 additional measures.

There were minimal differences in baseline accuracy scores among hospitals characterized by urban or rural location and small or large capacity,[35] but variation across hospitals served by different data vendors was more substantial. Rural hospitals and smaller hospitals generally received accuracy scores similar to those of urban hospitals and larger hospitals.[36] Among the hospitals that used JCAHO-certified data vendors to submit their quality data to the clinical warehouse, a higher percentage of

---

[34]For our analysis of baseline accuracy, the expanded-measure set includes the seven additional quality measures beyond the APU-measure set that HQA adopted for discharges after March 31, 2004. We found that some hospitals submitted data on the additional measures to the clinical warehouse for discharges occurring before that date, possibly because the hospitals were already collecting those data for JCAHO.

[35]We assessed hospital capacity in terms of the number of patient beds.

[36]For more detailed information on the relation of data accuracy to hospital characteristics and use of data vendors, see the tables in appendix III.

**GAO-06-54  Reliability of Hospital Quality Data**

hospitals served by certain data vendors met the 80 percent threshold than did the hospitals served by other data vendors (see app. III, table 8).[37]

## Passing the 80 Percent Threshold Is Statistically Uncertain for One-Fourth to One-Third of Hospitals

While the baseline level of data accuracy achieved by hospitals in the aggregate was well above the 80 percent threshold, for approximately one-fourth to one-third of hospitals the determination that a particular hospital met the 80 percent threshold was statistically uncertain. This uncertainty stems primarily from the small number of cases examined for accuracy from each hospital. Because CDAC's reabstraction of the data is limited to five patient records per quarter, the greater sampling variability found in small samples leads to relatively large confidence intervals, reflecting low statistical precision, for the accuracy score of any specific hospital.[38] Across all hospitals, the median difference between the upper and lower limits of the confidence interval was 14.0 percentage points using the APU-measure set for first-quarter discharges, dropping to 11.8 percentage points in the second quarter.[39] For the expanded-measure set, the median confidence interval was 14.6 percentage points in the first quarter and 13.0 percentage points in the second.

---

[37]The data that we obtained from CMS specifically identified data vendors that JCAHO had certified for its own performance reporting system. These data vendors submitted data to the clinical warehouse for 78 to 79 percent of the hospitals we analyzed for the two baseline quarters, while another 13 to 14 percent of hospitals directly submitted their own data.

[38]Statistical uncertainty occurs because different samples generally produce different results, due to variation among the individual patients selected for different samples. With larger samples, differences in the results obtained from one sample to another decrease. Calculating a confidence interval provides a way to assess the effect of sample variation on the results obtained. Confidence intervals are usually computed at the 95 percent level. So if 100 samples were selected, the result produced by 95 of them would likely fall between the low and high ends of the confidence interval. For example, one 300-plus-bed hospital in Virginia had an accuracy score of 83.3 for the second calendar quarter of 2004 using the expanded-measure set, with a confidence interval that ranged from 76.8 to 89.9. There is a 95 percent likelihood that any sample selected for that hospital would generate an accuracy score that was greater than 76 and lower than 90.

[39]The formula used to generate these confidence intervals takes into account variation in the number of individual data elements that were available in the five selected cases to compare the hospital's and CDAC's results. This is the same formula that is used by CMS, with one modification. Whereas CMS applied a one-tailed test at a 95 percent significance level to protect against hospitals receiving a failing score due to sampling error, we applied a two-tailed test at the 95 percent significance level to identify both failing and passing scores that were statistically uncertain. (See app. I.)

The wide confidence intervals meant that for a substantial number of hospitals it was statistically uncertain whether a different sample of cases would have altered their result from passing the 80 percent threshold to failing, or vice versa.[40] For most hospitals there was statistical certainty that their baseline accuracy score met CMS's 80 percent accuracy threshold. However, other hospitals had confidence intervals for their accuracy scores where the upper limit was 80 or above and the lower limit was less than 80. Because the confidence interval around the accuracy score computed for each of these hospitals bracketed the accuracy threshold set by CMS, their results were statistically uncertain.[41] Consequently, for these hospitals, the small number of cases that CMS examined was not sufficient to establish whether the hospital met the threshold level for data accuracy. One-third of all the hospitals that CMS assessed for accuracy fell into this uncertain category for first-quarter 2004 discharges using the APU-measure set. (See fig. 3.) This proportion declined to about one-fourth of the hospitals for the second quarter. When the expanded-measure set was used—as CMS has done when calculating its quarterly accuracy scores—the proportion of hospitals whose accuracy scores were statistically uncertain increased compared to the APU-measure set for both the first and the second quarter.

---

[40]Most, but not all, of the hospitals with statistically uncertain results had accuracy scores of 80 or above. See table 10 in appendix III.

[41]For example, if a hospital had a confidence interval that ranged from 77 to 90, taking multiple samples would lead to some samples generating accuracy scores at or above 80 and other samples generating scores of less than 80. Whether that hospital passed the 80 percent accuracy threshold would depend on which of those samples was actually selected.

**Figure 3: Percentage of Hospitals Whose Baseline Accuracy Score Confidence Intervals Clearly Exceed, Fall Below, or Include the 80 Percent Threshold, by Measure Set and Quarter**



January-March 2004 discharges

1.5%
32.2% 66.3%
The APU-measure set

1.9%
39.2% 58.9%
The expanded-measure set

April-June 2004 discharges

1.1%
26.3% 72.6%
The APU-measure set

1.2%
32.7% 66.1%
The expanded-measure set

Percentage of hospitals whose confidence interval clearly exceeds 80%

Percentage of hospitals whose confidence interval includes the 80% threshold

Percentage of hospitals whose confidence interval clearly falls below 80%
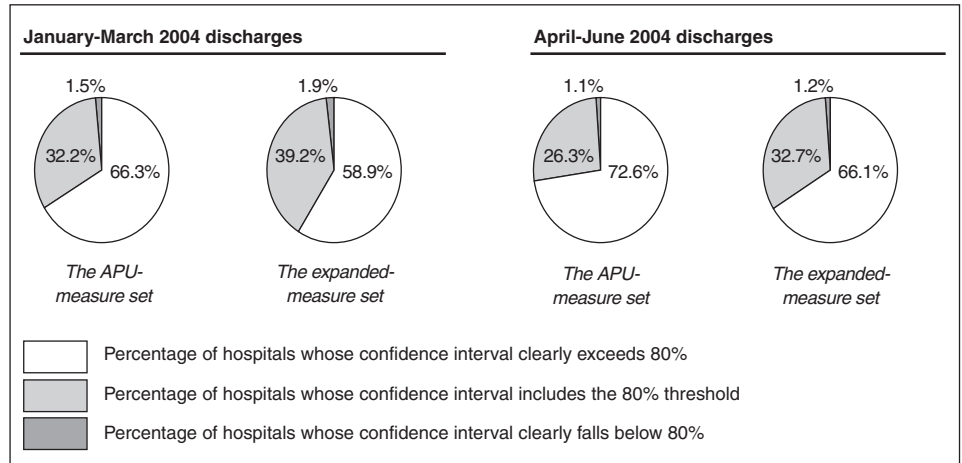
Source: GAO analysis of CMS data.

Note: The confidence interval is based on a 95 percent significance level. Calculation of the accuracy scores and confidence intervals for the expanded-measure set was based on all the measures for which a hospital submitted data, which could range from the APU measures alone to a maximum of 17—the APU measures plus as many as 7 additional measures.

These confidence intervals would narrow if CMS drew on multiple quarters of data to bring more cases into the computation of the accuracy scores. CMS has stated its intention to base this accuracy assessment on four quarters of hospital quality data, but so far every accuracy score it has generated and reported to hospitals has been based on a single quarter of data. Moreover, its implementation of the fiscal year 2006 payment update called for using only one quarter of data, with the possibility of adding one more quarter of data for hospitals that failed to meet the accuracy threshold based on the single quarter of data.[42]

---

[42]See 70 *Fed. Reg.* at 47422.

## No Data Were Available to Provide Baseline Assessment of Completeness of Hospital Quality Data

There were no data available from which to estimate a baseline level of completeness for the first two calendar quarters of data submitted for the APU program. In contrast to the system of quarterly reabstractions performed by CDAC to check the accuracy of quality data submitted by hospitals, CMS did not conduct any corresponding assessment of the extent to which all hospitals submitted data on all the cases, or a representative sample of such cases, that met CMS's eligibility criteria for the first two calendar quarters of 2004.

The information that CMS did collect was not suitable for estimating the baseline level of data completeness. The Medicare claims data analysis conducted by CMS on the first calendar quarter of data submitted for the APU program was not designed to provide valid information on the magnitude of data incompleteness for each hospital, which is what is needed to estimate a baseline level of data completeness. Although CMS could identify instances where certain hospitals failed to provide quality data on all eligible cases, CMS's analysis did not produce comparable information on data completeness for every hospital. As noted above, it lacked information on non-Medicare patients and could not adjust properly for the sample sizes that JCAHO-accredited hospitals would have drawn if they followed JCAHO's sampling rules rather than CMS's. The limitations in the CMS analysis would affect some hospitals more than others, depending on how many non-Medicare patients a hospital treated and whether it applied the JCAHO sampling rules. Consequently, had we used information from this analysis to estimate baseline data completeness, our results would have been distorted by the uneven impact of those factors on the information produced for different hospitals.[43]

In addition, we found no data for assessing the baseline completeness of the quality data provided by hospitals submitting samples of their eligible cases to the clinical warehouse. For hospitals that submitted a sample, their quality data could be incomplete, even if they submitted the expected number of cases, if their samples were not selected in a way that ensured they were representative of all a hospital's patients. If a hospital did not follow appropriate procedures to provide for random selection, the sample might not be representative and therefore could be incomplete. Because the available information from CMS focused on the number of cases submitted, and not on how they were selected, we could not address this aspect of data completeness.

---

[43]See appendix I for a more detailed description of this assessment.

## Other Reporting Systems Use Various Methods to Ensure Data Accuracy and Completeness, Notably an Independent Audit

Other reporting systems that collect clinical performance data have adopted various methods to ensure data accuracy and completeness, and officials from these systems stressed the importance of including an independent audit in these activities. Most other reporting systems that conduct independent audits incorporate three methods as part of their process that CMS does not use in its independent audit. Specifically, these systems include an on-site visit, focus their audit on a selected number of facilities or reporting entities, and review a minimum of 50 patient medical records per reporting entity.

### Other Reporting Systems Use Various Methods to Check Data

Other reporting systems that collect clinical performance data have adopted various methods to ensure data accuracy and completeness. To check data accuracy, all the other reporting systems we examined assess the data when they are submitted, typically using computers to detect missing or out-of-range data. (See app. II, tables 4 and 5.) In addition, all the other systems have developed standardized data collection processes and measures. When checking data completeness, all the other systems compare submitted data with data from another source, whether inside the facility, such as pharmacy or laboratory records, or outside the facility, such as state hospital discharge data or Medicare claims data. Officials reported that these analyses were done annually or had been done one time, and one said that additional studies were planned.[44] Officials from these systems also cite various other methods to consider when ensuring data accuracy and completeness, including reviewing established measures annually, identifying a point person at each facility to provide

---

[44]For example, on-site auditors from one reporting system compare the data submitted against catheterization laboratory schedules and hospital billing records for the previous 12 months. Another reporting system hired a contractor to perform a one-time study comparing patient assessment data submitted by a facility against its total Medicare claims to identify instances where patient assessments were missing.

consistency, establishing channels for ongoing communication, and providing training on a continuous basis.[45]

## Other Reporting Systems Conduct Independent Audits

Most other reporting system officials we interviewed conduct independent audits that include a comparison of submitted data to medical records. Most other reporting systems that conduct independent audits incorporate three methods as part of their process that CMS does not use in its independent audit. Specifically, they (1) include an on-site visit as part of their independent audit, (2) focus their audits on a selected number of facilities or reporting entities, and (3) review a minimum of 50 patient medical records per reporting entity during the auditing process. During an on-site visit, auditors are able to review patient medical records for accuracy and interview staff when additional information is needed. Auditors are also able to check the data submitted to their system against other data sources at the facilities, including physician notes, patient or resident rosters, billing records, laboratory records, and pharmacy records. In addition, because auditors from other reporting systems may not visit every facility,[46] the systems use various methods to focus the auditing process when selecting which facilities to visit. These include auditing a percentage of all eligible facilities, auditing facilities that did particularly well or poorly, and auditing a subset of facilities each year. Furthermore, most of the other reporting systems that conduct independent audits review a minimum of 50 patient medical records per audited entity as part of their independent auditing process. When selecting which patient medical records to review, some systems take a random sample of the patient population, one system reviews all deaths at the selected facility, and another reviews all instances where the patient died from shock as a result of percutaneous coronary intervention.

---

[45]We have also published a document that describes a flexible framework for assessing data reliability, including both accuracy and completeness, when assessing computer-processed data. This document offers procedures that can be adapted to varying circumstances. These procedures include conducting electronic data testing, such as logic tests; ensuring internal control systems are in place that check the data when they are entered into the system and limit access to the system; checking for missing data elements as well as missing case records; and reviewing related documentation, which may include tracing a sample of records large enough to estimate an error rate back to their source documents. See GAO, *Assessing the Reliability of Computer-Processed Data*, GAO-03-273G (Washington, D.C.: October 2002) External Version 1.

[46]An official from one reporting system said that budgetary constraints limit the number of on-site audits that the system can perform. As a result, auditors from that system focus their review on hospitals with outcomes that fall above and below the systemwide average.

Officials at other reporting systems we interviewed and an expert in the field stressed the importance of the independent audit. For example, an official from one of the other reporting systems said that audits conducted by an independent third party are "the best way" to ensure data accuracy and completeness. An official from another reporting system said that having someone independently check the data is "one of the most important things" that an organization can do to check data accuracy and completeness. Additionally, an expert we interviewed said that independent, external audits are "essential." Though most of the other reporting systems employ an independent auditing process, officials from one system that has yet to implement such a process said their organization recognizes the importance of independently checking the data and is currently designing and implementing an independent auditing process.

## Conclusions

Data collected for the APU program affect the payment received by hospitals from Medicare and are used to inform the public about hospital quality. For both these purposes, it is important that CMS is able to ensure that the data are reliable in terms of both accuracy and completeness.

CMS has put in place an ongoing process for assessing the accuracy of quality data submitted by hospitals, but the process has limitations. Although CMS checks the accuracy of data electronically as they are submitted and through an independent audit conducted by CDAC, the latter process is limited by the selection of only five cases per quarter per hospital, regardless of the hospital's size. Most hospitals had high baseline accuracy scores that were statistically certain. However, for about one-fourth to one-third of all the hospitals that CMS assessed for the first two calendar quarters of 2004, CMS's determination as to whether the hospital met its accuracy standard was statistically uncertain. This was due primarily to the small number of cases selected for an audit. Although CMS has stated its intention to look at more cases by pooling reabstraction results from more than one calendar quarter, all of the hospital accuracy reports that it has generated to date have been based on a single quarter of data. Officials from other reporting systems that collect clinical performance data told us that they also use an independent audit to check data accuracy, but generally sample a larger number of patient medical records, either by sampling a percentage of total cases submitted or by identifying a minimum number of cases in the sample. In addition, most other reporting systems focused their audits on a selected number of facilities.

In contrast to CMS's establishment of an ongoing process for assessing data accuracy, the agency has not put in place an ongoing process to check the completeness of the data that hospitals submit. Because of the purposes for which these data may be used, there could be an incentive for hospitals to selectively report data on cases that score well on the quality measures. With no ongoing way to check completeness, CMS does not know whether or how often hospitals submit incomplete data. We believe this is a significant gap in oversight. The process used for the fiscal year 2005 annual payment update compared hospital submissions to Medicare claims data, but as CMS has noted, this did not provide a comparable assessment of each hospital's data, even for Medicare patients alone. Moreover, in its comparison of hospital quality data submissions with Medicare claims for the fiscal year 2006 update, CMS identified more than 100 hospitals that had treated eligible patients in a given quarter but had not submitted data on a single case for that quarter to the clinical warehouse. Yet CMS has not asked hospitals to certify that the data they have submitted constitute all, or a representative sample, of the eligible patient population. The various methods used by other reporting systems to check the completeness of data illustrate the variety of approaches that are available. These include conducting on-site visits as part of their independent audit, comparing data submissions to data from another source maintained by the facility or external to it, and performing such checks annually or planned at specified intervals.

Given CMS's plans to continue public reporting efforts after the APU program ends, we believe that processes for checking the reliability of data should continue to be refined in order for the individuals and organizations that use the data to have confidence in the information.

## Recommendations for Executive Action

In order for CMS to help ensure the reliability of the quality data it uses to produce information on hospital performance, we recommend that the CMS Administrator undertake the following three actions:

- focusing on the subset of hospitals for which it is statistically uncertain if they met CMS's accuracy threshold in one or more previous quarters, increase the number of patient records reabstracted by CDAC in a subsequent quarter so that the proportion of hospitals with statistically uncertain results is reduced;
- require hospitals to certify that they took steps to ensure that they submitted data on all eligible patients, or a representative sample thereof; and

- assess the level of incomplete data submitted by hospitals for the APU program to determine the magnitude of underreporting, if any, in order to refine how completeness assessments may be done in future reporting efforts.

## Agency Comments

In commenting on a draft of this report, CMS stated it appreciated our analysis and recommendations. (CMS's comments appear in app. IV.) The agency noted that the APU program led to a dramatic increase in the number of hospitals that submitted data on the designated 10 quality measures, resulting in public reporting of quality data for about 3,600 hospitals on the agency's Web site. In addition, CMS described the steps it had taken to ensure the accuracy and completeness of the quality data submitted by hospitals for the APU program. It said that the methods it had used were sound, but it agreed that the quality and completeness of the data must be improved.

With respect to reducing the statistical uncertainty of its assessments of the accuracy of hospital quality data submissions, CMS agreed that the quarterly accuracy assessments based on five patient charts can have considerable sampling error and stated that it would improve the stability of its accuracy assessments by using data from four calendar quarters when it assessed hospital eligibility for the fiscal year 2007 annual payment update. CMS stated a concern with having sufficient time within the current data submission schedule to increase the number of patient records reabstracted. However, we recommended in the draft report that hospitals with statistically uncertain results in one or more previous quarters have an increased number of records reabstracted. The assessment of statistical uncertainty for a hospital and the reabstraction of additional records do not need to occur within the same quarter. We have modified slightly the wording of the recommendation to clarify the intended timing of these additional reabstractions.

With respect to ensuring the completeness of quality data submitted by hospitals, CMS agreed that it needs to improve its methods. CMS noted that its comparison of hospital data quality submissions to the claims filed by those hospitals to be paid for treating Medicare beneficiaries uncovered numerous discrepancies. The agency agreed with our recommendation to require hospitals to formally attest to the completeness of the quality data that they submit quarterly. In addition, CMS stated that it would also require each hospital to report the total number of Medicare and non-Medicare patients who were eligible for quality assessment under the APU program.

In terms of assessing the level of incomplete data for the APU program, CMS said it had a process in place to accomplish this, but as we stated in the draft report, CMS's process did not cover all patients and all hospitals because it lacked information on non-Medicare patients even though hospitals were required to submit data on both Medicare and non-Medicare patients. Additionally, the tests that CMS applied could detect incomplete data for only a limited subset of hospitals, in contrast to its assessment of data accuracy which covered all hospitals that submitted data on six or more cases in a quarter. CMS acknowledged it could assess completeness only for Medicare patients, but said that by requiring hospitals to report an aggregate count of all eligible patients, it would henceforth have the data needed to assess the completeness of both Medicare and non-Medicare quality data submissions. The agency stated it will use these data to provide quarterly feedback to hospitals about the accuracy and completeness of their data submissions, and require them to explain discrepancies between the data they have submitted for the APU program and the aggregate count of eligible patients they have reported. CMS has not said that it will determine the magnitude of underreporting for the program as a whole, as we recommended. Additionally, by relying on the hospitals themselves to supply data on the number of non-Medicare patients, CMS's proposed approach lacks an independent verification of the completeness of submitted data. This contrasts with the practice of most of the other reporting systems we contacted, as well as experts in the field, who generally underscored the importance of independently checking both the accuracy and the completeness of the quality data.

As arranged with your offices, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days after its issue date. At that time, we will send copies of this report to the Administrator of CMS and other interested parties. We will also make copies available to others on request. In addition, the report will be available at no charge on GAO's Web site at http://www.gao.gov.

If you or your staffs have any questions about this report, please contact me at (202) 512-7101 or BascettaC@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made major contributions to this report are listed in appendix V.

*Cynthia Bascetta*

Cynthia A. Bascetta
Director, Health Care

# Appendix I: Scope and Methodology

To determine the processes used by the Centers for Medicare & Medicaid Services (CMS) to ensure the accuracy and completeness of data submitted by hospitals for the Annual Payment Update program (APU program), we interviewed both CMS officials and staff at DynKePRO—which operates the Clinical Data Abstraction Center (CDAC)—and the Iowa Foundation for Medical Care (IFMC), two contractors that perform data collection and data quality monitoring tasks for the APU program. In addition, we reviewed documentation on the program available publicly on the Quality Net Exchange Web site[1] and the Web sites of several quality improvement organizations (QIO)—contractors to CMS that provide technical assistance to hospitals on the APU program—as well as documents on the APU program provided to us at our request by CMS. We also obtained access to CMS's intranet system and searched for relevant memorandums and other documents regarding CMS's policies and requirements for hospitals that participated in the APU program. To gain insights from other groups involved in the APU program, we interviewed officials from two or more QIOs, state hospital associations, and hospital data vendors that submitted data to the IFMC-operated database for their hospital clients.

Our assessment of the baseline accuracy of the initial APU program data depended on the availability of suitable information from CMS. We examined CMS's reabstraction process to determine if the CDAC assessments of data accuracy would be appropriate for that purpose. Reabstraction is the re-collection of clinical data for the purpose of assessing the accuracy of data abstractions performed by hospitals. In the APU program, CDAC compares data reported by the hospitals to those it has independently obtained from the same medical records. CDAC has instituted a range of procedures, including training of its abstractors and continuous monitoring of interrater reliability, intended to ensure that its abstractors understand and follow its detailed guidance for arriving at abstraction determinations that are correct in terms of CMS's data specifications. We interviewed CDAC staff and observed the implementation of these procedures during a site visit at the CDAC facility. On the basis of this information we concluded that it would be appropriate for us to use the results of the CDAC reabstractions to estimate baseline data accuracy for the APU program.

---

[1]We downloaded various documents from the www.qnetexchange.org Web site between December 21, 2004, and January 10, 2006.

We obtained the results of the reabstractions that CDAC had conducted on samples of the patients for whom hospitals had submitted data from the first two quarters of 2004. These two quarters were the first two data submissions made by hospitals under the APU program and the most recent available when we conducted these analyses. They constituted 20,465 patient records for the first quarter and 20,259 for the second. These files showed, for each data element that CMS used in assessing abstraction accuracy, the correct entry as determined by the CDAC abstractors and whether this matched the value that the hospital had reported. We applied CMS's algorithms for computing hospital scores on the expanded-measure set in order to determine the extent of missing or invalid data. We found that approximately 2 to 3 percent of patient records could not be scored on any given APU measure due to missing data. We excluded from the analysis records from critical access hospitals and acute care hospitals in Maryland and Puerto Rico (which are paid under different payment systems than other acute care hospitals and therefore are not subject to a reduced annual payment update under the APU program[2]) and a small number of records not related to the three medical conditions covered by the APU program.[3]

Next we applied the scoring rules developed by CMS to assess the accuracy of hospital abstractions. We calculated the accuracy score for each hospital in each quarter, using the data elements needed for the APU-measure set and, separately, for the expanded-measure set. Accuracy scores for the expanded-measure set are based on all the measures for which a hospital submitted data, which could range from the APU measures alone to a maximum of 17—the 10 measures in the APU-measure set plus the 7 additional measures adopted by the Hospital Quality Alliance for hospital discharges through the second calendar quarter of 2004. These scores represented the proportion of data elements where CDAC and the hospital agreed, summing across all the assessed data elements for the five sampled cases. We then calculated the distribution of those scores, and the proportion of hospitals that met or

---

[2]CMS included hospitals in Puerto Rico in its list of hospitals qualifying for the full fiscal year 2005 update, but determined in conjunction with the fiscal year 2006 payment update decision that Puerto Rico's hospitals were exempt from the APU program requirements. Hospitals in Puerto Rico receive prospective payments from Medicare, but under a different system than other hospitals.

[3]The records we excluded were 536 surgery cases for the first quarter and 604 surgery cases for the second quarter, from hospitals providing data on surgical infection prevention measures.

exceeded the 80 accuracy threshold that CMS had set. Next we calculated the confidence interval for each of those accuracy scores, using the formula that CMS had selected for that purpose. However, whereas CMS applied a one-tailed test—passing any hospital that had a confidence interval whose upper bound reached 80 or above—we applied a two-tailed test to assess the statistical uncertainty attached to both passing and failing the threshold. The one-tailed test that CMS applied prevented hospitals from losing their full annual payment update on the basis of their accuracy score if there was less than a 95 percent probability that a score below 80 would have remained below 80 in another sample. This meant that hospitals with large confidence intervals could have accuracy scores well below 80 and still pass the CMS accuracy requirement. Our analysis focused instead on assessing the level of statistical certainty for all the accuracy scores, both above and below the 80 percent threshold. We sought to identify passing as well as failing scores that could have changed with another sample. To do so, we applied a two-tailed test and observed whether a hospital's confidence interval bracketed the 80 percent threshold.

To provide descriptive information about variation in the accuracy scores obtained by hospitals in different situations, we collected additional information about the hospitals from other sources. From the Medicare Provider of Services file we obtained the Social Security Administration metropolitan statistical area code (referred to as the SSA MSA code) and Social Security Administration metropolitan statistical area size code (referred to as the SSA MSA size code) to distinguish between urban and rural hospitals. We also obtained from that source the total number of Medicare-certified beds in order to categorize hospitals by size. To compare the accuracy scores of hospitals that employed different data vendors, we obtained from IFMC the identification codes (but not the names) of the various data vendors certified by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) that had submitted to the clinical warehouse data for the APU program on behalf of hospitals they served. Those codes were also available in the case tracking information for the patient records in the CDAC database. We then identified for each CDAC reabstraction whether the case had originally been submitted by a JCAHO-certified data vendor, and if so, which one. These data were aggregated to generate accuracy scores for each hospital that consistently submitted its quality data through one data vendor in a given quarter. This allowed us to determine the proportion of hospitals served by each JCAHO data vendor that met CMS's 80 percent accuracy threshold. We also calculated the proportion of hospitals that submitted their own quality data to CMS (identified in the CDAC case tracking

information by the hospital's Medicare provider ID number) that met the accuracy threshold. Although this analysis was limited to data vendors that were JCAHO-certified, those vendors collectively submitted data to the clinical warehouse for 78 to 79 percent of the hospitals we analyzed in the two baseline quarters. Another 13 to 14 percent of hospitals directly submitted their own data, and we do not have information on how the remaining hospitals submitted data to the clinical warehouse.

As was the case for our baseline accuracy assessment, our assessment of the baseline completeness of the data submitted for the APU program depended on the availability of suitable data from CMS. Specifically, we considered using CMS's estimates of minimum expected cases derived from Medicare claims data to arrive at estimates of baseline completeness. The CMS officials we spoke with noted that there were numerous reasons why the two data sources—quality data submissions for the APU program and cases billed to Medicare—would be expected to diverge, apart from any underreporting of quality data by hospitals. The claims data were limited to Medicare fee-for-service patients, whereas the hospitals were obliged to submit quality data on all patients over 18 years of age (over 28 days old for most pneumonia measures), including patients belonging to Medicare health maintenance organizations. In addition, hospitals with large numbers of cases could draw samples for the quality data, but would bill for all patients. In making adjustments to its number of "expected cases" to take account of sampling, CMS found that it could not reliably identify the hospitals that should have followed the JCAHO sampling rules, which would result in larger-sized samples. Therefore, in calculating the number of cases it expected hospitals to have submitted to the clinical warehouse, CMS applied to all hospitals across the board the expectation of smaller samples based on rules that pertained to hospitals not accredited by JCAHO. Finally, the Medicare data used for the comparison was an average volume recorded over the previous 2 years, not claims filed for the quarter to which the quality data applied.

We found that these limitations precluded our using information from CMS's Medicare claims analysis to assess the baseline completeness of the data submitted by hospitals for the APU program. CMS's comparison of hospital quality data submissions to the clinical warehouse to its estimated number of "expected cases" might have served CMS's purposes, by identifying at least some instances of significant discrepancy between the number of cases for which quality data were submitted and claims filed. However, we determined that it would not provide a reasonable estimate of the magnitude of data completeness for all hospitals. Because the limitations in the CMS analysis would affect some hospitals more than

others, depending on how many non-Medicare patients a hospital treated and whether it applied the JCAHO sampling rules, we concluded that using information from this analysis to estimate baseline data completeness would lead to results that were distorted by the uneven impact of those factors on the information produced for different hospitals.

To obtain information on other processes that could be used to check data accuracy and completeness, we interviewed officials from organizations that administer reporting systems that collect clinical performance data. To select these organizations, we took several steps. We reviewed reports on reporting systems, including two issued by QIOs: IPRO's *2003 Review of Hospital Quality Reports* and Delmarva Foundation's *The State-of-the-Art of Online Hospital Public Reporting: A Review of Forty-Seven Websites*.[4] We solicited input from the authors of each report and interviewed academic researchers who have researched methods of assessing the reliability of performance data. We used on-line resources to obtain information on federal- and state-administered surveillance efforts. Our selection criteria focused on systems that collected clinical data, as opposed to administrative or claims data, and that were mentioned most often in the reports and interviews cited above. To ensure variation, we selected a mix of systems, including those run by public and private organizations, those receiving data from hospitals and those receiving data from other types of providers, and those collecting data across a range of medical conditions and those collecting data on specific medical conditions. Using a structured protocol, we interviewed officials from the following organizations: JCAHO, National Committee for Quality Assurance, Society of Thoracic Surgeons, California Office of Statewide Health Planning and Development, New York State Department of Health, CMS (the units responsible for monitoring nursing home care regarding the Data Assessment and Verification Project (DAVE) contract), and the American College of Cardiology. Each organization reviewed and confirmed the accuracy of the information presented in appendix II.

Our analysis is based on the quality measures established for the APU program and the information available as of September 2005 on the accuracy and completeness of data submitted by hospitals for that

---

[4]IPRO, *2003 Review of Hospital Quality Reports for Health Care Consumers, Purchasers and Providers* (Lake Success, N.Y.: October 2003); Delmarva Foundation and the Joint Commission on Accreditation of Healthcare Organizations, *The State-of-the-Art of Online Hospital Public Reporting: A Review of Forty-Seven Websites* (Easton, Md.: September 2004).

program. We did not evaluate the appropriateness of these quality measures relative to others that could have been selected. Nor did we examine the actual performance by hospitals on the measures (e.g., how often they provide a particular service or treatment). Our analysis of the baseline level of accuracy and completeness of data submitted for the APU program is based on the procedures developed by CMS to validate the data submitted. We have not independently compared the data submitted by hospitals to the original patient clinical records.

We conducted our work from November 2004 through January 2006 in accordance with generally accepted government auditing standards.

**Table 3: Background Information on CMS and Other Reporting Systems**

| | Centers for Medicare & Medicaid Services (CMS) | Other reporting systems | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | American College of Cardiology (ACC) | California Office of Statewide Health Planning and Development | Data Assessment and Verification Project (DAVE)[a] | Joint Commission on Accreditation of Healthcare Organizations (JCAHO)[b] | National Committee for Quality Assurance (NCQA) | New York State Department of Health | Society of Thoracic Surgeons (STS) |
| **Organization status** | Public | Private, nonprofit | Public | Public | Private, nonprofit | Private, nonprofit | Public | Private, nonprofit |
| **Data submitted by** | Hospitals paid under the Inpatient Prospective Payment System | Facilities with at least one catheterization laboratory (includes in-hospital, freestanding, and/or mobile catheterization laboratories) | Hospitals where cardiac surgeries are performed | Nursing homes | JCAHO-accredited hospitals | Health plans | Hospitals that perform cardiac surgery and/or percutaneous coronary intervention (PCI) | Hospitals, surgeons |
| **Reporting requirement** | [c] | Voluntary[d] | Mandatory | Mandatory | Mandatory[e] | Mandatory[e] | Mandatory | Voluntary |
| **Are the data publicly reported?** | Yes | No | Yes | Yes | Yes | Yes[f] | Yes | No |
| **Types of conditions for which data are submitted** | Cardiac - acute myocardial infarction (AMI), heart failure (HF)<br><br>Pneumonia | Cardiac - diagnostic cardiac catheterization, PCI | Cardiac – coronary artery bypass grafting (CABG) | Resident health care<br><br>Resident health status | Cardiac - AMI, HF<br><br>Pneumonia<br><br>Pregnancy<br><br>Surgical infection prevention | Preventive care, acute and chronic conditions | Cardiac - CABG, PCI, and valve surgery | Cardiac - CABG, aortic and mitral valve<br><br>General thoracic surgery<br><br>Congenital heart surgery |
| **Number of facilities reporting** | 3,839[g] | 611[h] | 120 | 16,266[i] | ~3,350 | 560 | 49 | 700 |
| **Approximate program duration** | 2 years | 7 years | 2 years[j] | 1 year | 3 years | 14 years | 16 years | 16 years |

Sources: CMS, ACC, California Office of Statewide Health Planning and Development, JCAHO, NCQA, New York State Department of Health, and STS.

[a]DAVE is a CMS contract to assess the reliability of minimum data set assessment data that are submitted by nursing homes. Minimum data set assessments are a minimum data set of core elements to use in conducting comprehensive assessments of patient conditions and care needs. These assessments are collected for all residents in nursing homes that serve Medicare and Medicaid beneficiaries.

[b]JCAHO provided information about its ORYX® initiative, which integrates outcome and other performance measurement data into the accreditation process.

[c]Under Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, hospitals shall submit data for a set of indicators established by the Department of Health and Human Services (HHS) as of November 1, 2003, related to the quality of inpatient care. Section 501 (b) also provides that any hospital that does not submit data on the 10 quality measures specified by the Secretary of Health and Human Services will have its annual payment update reduced by 0.4 percentage points for each fiscal year from 2005 through 2007.

[d]Some states and insurance companies have started to require hospital participation.

[e]Data submission is mandatory to maintain accreditation.

[f]Only audited data are publicly reported.

[g]The number of hospitals that submitted data to receive their annual payment update for fiscal year 2005.

[h]The number of facilities enrolled in ACC's National Cardiovascular Data Registry® as of July 13, 2005.

[i]This number represents the number of nursing homes that submitted minimum data set assessments between January 1, 2004, and December 31, 2004. Accuracy estimates are made by selecting a random sample of records for off-site and on-site medical record review.

[j]Mandatory reporting of performance data began in 2003.

**Table 4: Processes Used by CMS and Other Reporting Systems to Ensure Data Accuracy**

| | Centers for Medicare & Medicaid Services (CMS) | Other reporting systems | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | American College of Cardiology (ACC) | California Office of Statewide Health Planning and Development | Data Assessment and Verification Project (DAVE)[a] | Joint Commission on Accreditation of Healthcare Organizations (JCAHO)[b] | National Committee for Quality Assurance (NCQA) | New York State Department of Health | Society of Thoracic Surgeons (STS) |
| **Processes** | | | | | | | | |
| Training | ● | ● | ● | ● | ● | ● | ● | ● |
| Standardized measures or definitions | ●[c] | ● | ● | ● | ●[c] | ● | ● | ● |
| Standardized processes for data collection | ● | ● | ● | ● | ● | ● | ● | ● |
| Automated data edits when the data come in as part of the data quality assurance process (identify missing or out-of-range data) | ● | ● | ● | ●[d] | ● | ● | ● | ● |
| Independent audits | ● | ● | ● | ● | [e] | ● | ● | [f] |
| On-site audits | | ● | ● | ● | | ● | ● | [f] |
| Medical record review | ● | ● | ● | ● | | ● | ● | [f] |
| **Sample size** | | | | | | | | |
| Patients | 5 records | 10% random sample of medical records, 50 record minimum[g] | 70 records | 13-16 records[h] | Not applicable | 60 records | 50 records[i] | [j] |
| Facilities | All | 10% random sample of eligible sites[k] | Outliers and near-outliers for mortality | 69 | Not applicable | All | 20 programs/year[l] | [m] |

Sources: CMS, ACC, California Office of Statewide Health Planning and Development, JCAHO, NCQA, New York State Department of Health, and STS.

[a]DAVE is a CMS contract to assess the reliability of minimum data set assessment data that are submitted by nursing homes. Minimum data set assessments are a minimum data set of core elements to use in conducting comprehensive assessments of patient conditions and care needs. These assessments are collected for all residents in nursing homes that serve Medicare and Medicaid beneficiaries.

[b]JCAHO provided information about its ORYX® initiative, which integrates outcome and other performance measurement data into the accreditation process.

[c]CMS and JCAHO have worked to align their measures. A common set of measures took effect for discharges occurring on or after January 1, 2005.

[d]Data checks occur at the state level, for example, the state health department, before the data are accessed by DAVE.

[e]JCAHO performs independent audits of data vendors.

[f]STS is planning to incorporate an independent audit into its system. STS officials plan on including an on-side audit and medical record review as part of their audit system.

[g]The 10 percent random sample of medical records is based on annual percutaneous coronary intervention volume.

[h]The number of cases and facilities identified are limited to on-site audits. Additional cases are reviewed as part of the off-site medical record review process.

[i]Auditors review 100 percent of records when significant discrepancies are identified between the chart and what the hospital reported on specific risk factors. In addition, medical record documentation is reviewed for 100 percent of cases with the risk factors "shock" or "stent thrombosis".

[j]STS plans to review a minimum of 30 records as a part of its independent auditing process.

[k]ACC defines eligible sites as those facilities with a minimum of 50 records to be abstracted over a specified number of quarters.

[l]New York State Department of Health typically reviews 20 programs per year. In some instances that can mean percutaneous coronary intervention and cardiac surgery at the same hospital, which would count as two programs.

[m]STS plans on visiting 24 facilities per year as a part of its independent auditing process.

**Table 5: Processes Used by CMS and Other Reporting Systems to Ensure Data Completeness**

| | Centers for Medicare & Medicaid Services (CMS) | Other reporting systems | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | American College of Cardiology (ACC) | California Office of Statewide Health Planning and Development | Data Assessment and Verification Project (DAVE)[a] | Joint Commission on Accreditation of Healthcare Organizations (JCAHO)[b] | National Committee for Quality Assurance (NCQA) | New York State Department of Health | Society of Thoracic Surgeons (STS) |
| **Processes** | | | | | | | | |
| Training | | ● | ● | ● | | ● | ● | ● |
| Concurrent review[c] | | ● | | ● | | ● | | |
| Independent audits | | ● | ● | ● | [d] | ● | ● | [e] |
| On-site audits | | ● | ● | ● | | ● | ● | [e] |
| Comparison to another data source | ● | ● | ● | ● | ● | ● | ● | ● |
| **Data sources** | | | | | | | | |
| Billing | Medicare claims data | Hospital billing records | ICD-9 codes[f] | Medicare claims data | ICD-9 codes[f] | | Statewide planning and research cooperative system (SPARCS) data | Medicare provider analysis and review (MEDPAR) data |
| Other | | Patient medical records, catherization laboratory logs; physician notes | State death files | Resident rosters | | Pharmacy records, laboratory records | | |
| **Frequency of data completeness review** | Twice[g] | Annually[h] | Annually | Once[i] | Annually[j] | Annually | Annually | Once[k] |

Sources: CMS, ACC, California Office of Statewide Health Planning and Development, JCAHO, NCQA, New York State Department of Health, and STS.

[a]DAVE is a CMS contract to assess the reliability of minimum data set assessment data that are submitted by nursing homes. Minimum data set assessments are a minimum data set of core elements to use in conducting comprehensive assessments of patient conditions and care needs. These assessments are collected for all residents in nursing homes that serve Medicare and Medicaid beneficiaries.

[b]JCAHO provided information about its ORYX® initiative, which integrates outcome and other performance measurement data into the accreditation process.

[c]Under concurrent review, auditors assess data as they are being collected.

[d]JCAHO performs independent audits of data vendors.

[e]STS is planning to incorporate an independent audit into its system. STS officials plan on including an on-side audit as part of their audit system.

[f]The International Classification of Diseases, Ninth Revision (ICD-9) codes were designed to promote international comparability in the collection, processing, classification, and presentation of mortality statistics.

[g]CMS conducted two separate one-time studies that compared Medicare claims data to submitted data.

[h]Data completeness reviews are conducted annually for randomly selected sites as part of the on-site audit process and quarterly for data submissions.

[i]A one-time study was conducted; additional studies are planned.

[j]At a minimum, data completeness reviews are conducted annually.

[k]A one-time study was conducted.

# Appendix III: Data Tables on Hospital Accuracy Scores

Rural hospitals and smaller hospitals generally received accuracy scores that differed minimally from those of urban hospitals and larger hospitals. (See tables 6 and 7.) To the extent there are small differences across categories, they do not show a consistent pattern based on geographic location or size.

**Table 6: Median Hospital Baseline Accuracy Scores, by Hospital Characteristic, Quarter, and Measure Set**

| Hospital characteristic | January-March 2004 discharges | | April-June 2004 discharges | |
|---|---|---|---|---|
| | Median accuracy score for APU-measure set | Median accuracy score for expanded-measure set | Median accuracy score for APU-measure set | Median accuracy score for expanded-measure set |
| Urban | 92.7 | 90.0 | 94.2 | 91.5 |
| Rural | 93.0 | 91.1 | 93.8 | 91.7 |
| < 50 beds | 93.0 | 91.2 | 93.9 | 91.8 |
| 50-99 beds | 93.2 | 91.1 | 94.2 | 92.2 |
| 100-199 beds | 92.9 | 90.5 | 94.1 | 91.3 |
| 200-299 beds | 93.0 | 90.1 | 94.2 | 91.7 |
| 300-399 beds | 92.7 | 89.8 | 93.9 | 91.0 |
| 400-499 beds | 92.0 | 89.5 | 93.8 | 91.1 |
| 500+ beds | 92.0 | 89.0 | 94.1 | 91.0 |
| All hospitals | 92.9 | 90.4 | 94.1 | 91.6 |

Source: GAO analysis of CMS data.

Note: Calculation of accuracy scores for the expanded-measure set was based on all the measures for which a hospital submitted data, which could range from the APU measures alone to a maximum of 17—the APU measures plus as many as 7 additional measures.

**Table 7: Proportion of Hospitals with Baseline Accuracy Scores Not Meeting 80 Percent Threshold, by Hospital Characteristic, Quarter, and Measure Set**

| Hospital characteristic | January-March 2004 discharges | | April-June 2004 discharges | |
|---|---|---|---|---|
| | Percentage not meeting threshold for APU-measure set | Percentage not meeting threshold for expanded-measure set | Percentage not meeting threshold for APU-measure set | Percentage not meeting threshold for expanded-measure set |
| Urban | 10.3 | 14.4 | 7.7 | 10.3 |
| Rural | 9.1 | 11.6 | 8.9 | 9.6 |
| < 50 beds | 9.4 | 12.8 | 10.3 | 12.0 |
| 50-99 beds | 9.6 | 12.4 | 8.3 | 8.5 |
| 100-199 beds | 8.7 | 12.3 | 8.6 | 9.8 |
| 200-299 beds | 9.5 | 12.8 | 6.0 | 9.3 |
| 300-399 beds | 11.8 | 15.0 | 6.5 | 8.6 |
| 400-499 beds | 10.6 | 14.1 | 8.1 | 11.1 |
| 500+ beds | 12.2 | 16.6 | 8.6 | 12.2 |
| All hospitals | 9.8 | 13.2 | 8.2 | 10.0 |

Source: GAO analysis of CMS data.

Note: Calculation of accuracy scores for the expanded-measure set was based on all the measures for which a hospital submitted data, which could range from the APU measures alone to a maximum of 17—the APU measures plus as many as 7 additional measures. CMS deems hospitals that achieve an accuracy score of 80 or better as having met its requirement to submit accurate data.

Accuracy scores among hospitals whose data were submitted to CMS by different JCAHO-certified vendors varied more, especially in the percentage of the hospitals that failed to meet the 80 percent threshold. (See table 8.) Collectively, these data vendors submitted data to the clinical warehouse for approximately 78 to 79 percent of hospitals affected by the APU program in the two baseline quarters, while another 13 to 14 percent of hospitals directly submitted their own data. For large data vendors (serving more than 100 hospitals), medium vendors (serving between 20 and 100 hospitals), and small vendors (serving fewer than 20 hospitals), there was marked variation within each size grouping in the proportion of the vendors' hospitals that did not meet the accuracy threshold. Such variation could reflect differences in the hospitals served by different vendors as well as differences in the services provided by those vendors.

**Table 8: Percentage of Hospitals with Baseline Accuracy Scores Not Meeting 80 Percent Threshold, by JCAHO-Certified Vendor Grouped by Number of Hospitals Served, Quarter, and Measure Set**

| Vendors, grouped by number of hospitals served | Percentage not meeting threshold for APU-measure set | | Percentage not meeting threshold for expanded-measure set | |
|---|---|---|---|---|
| | January-March 2004 discharges | April-June 2004 discharges | January-March 2004 discharges | April-June 2004 discharges |
| **Large vendors** | | | | |
| Vendor 1 | 2.6 | 2.6 | 3.9 | 2.6 |
| Vendor 2 | 7.1 | 7.2 | 9.3 | 7.2 |
| Vendor 3 | 7.7 | 9.5 | 14.0 | 11.3 |
| Vendor 4 | 10.1 | 9.8 | 11.1 | 10.2 |
| Vendor 5 | 11.1 | 8.4 | 14.4 | 10.4 |
| Vendor 6 | 12.2 | 10.4 | 16.5 | 11.3 |
| Vendor 7 | 12.4 | 9.0 | 12.4 | 13.6 |
| Vendor 8 | 13.3 | 5.8 | 15.8 | 7.9 |
| **Medium vendors** | | | | |
| Vendor 9 | 2.4 | 4.5 | 2.4 | 2.3 |
| Vendor 10 | 3.4 | 3.1 | 3.4 | 6.3 |
| Vendor 11 | 4.2 | 6.8 | 6.9 | 6.8 |
| Vendor 12 | 4.8 | 4.8 | 4.8 | 6.5 |
| Vendor 13 | 4.9 | 2.8 | 4.9 | 2.8 |
| Vendor 14 | 6.4 | 4.3 | 8.5 | 6.4 |
| Vendor 15 | 7.1 | 6.0 | 7.1 | 7.5 |
| Vendor 16 | 7.6 | 5.0 | 19.0 | 13.8 |
| Vendor 17 | 7.9 | 2.6 | 9.2 | 2.6 |
| Vendor 18 | 8.0 | 3.4 | 12.0 | 6.9 |
| Vendor 19 | 8.8 | 2.9 | 26.5 | 8.8 |
| Vendor 20 | 12.1 | 5.5 | 17.6 | 7.7 |
| Vendor 21 | 13.5 | 5.6 | 13.5 | 8.3 |
| Vendor 22 | 15.2 | 13.9 | 17.7 | 17.7 |
| Vendor 23 | 18.4 | 10.0 | 28.6 | 12.0 |
| **Small vendors** | | | | |
| Vendor 24 | 0.0 | 11.8 | 0.0 | 11.8 |
| Vendor 25 | 0.0 | 7.1 | 0.0 | 7.1 |
| Vendor 26 | 0.0 | 0.0 | 0.0 | 0.0 |
| Vendor 27 | 0.0 | 16.7 | 0.0 | 16.7 |
| Vendor 28 | 0.0 | 0.0 | 0.0 | 0.0 |
| Vendor 29 | 0.0 | 0.0 | 0.0 | 0.0 |

| Vendors, grouped by number of hospitals served | Percentage not meeting threshold for APU-measure set | | Percentage not meeting threshold for expanded-measure set | |
| --- | --- | --- | --- | --- |
| | January-March 2004 discharges | April-June 2004 discharges | January-March 2004 discharges | April-June 2004 discharges |
| Vendor 30 | 0.0 | 0.0 | 0.0 | 0.0 |
| Vendor 31 | 8.3 | 0.0 | 16.7 | 0.0 |
| Vendor 32 | 9.1 | 8.3 | 9.1 | 16.7 |
| Vendor 33 | 9.1 | 0.0 | 27.3 | 0.0 |
| Vendor 34 | 10.0 | 9.1 | 10.0 | 9.1 |
| Vendor 35 | 11.1 | 11.1 | 11.1 | 11.1 |
| Vendor 36 | 20.0 | 33.3 | 60.0 | 33.3 |
| Vendor 37 | 33.3 | 0.0 | 33.3 | 0.0 |
| Vendor 38 | 33.3 | 0.0 | 33.3 | 0.0 |
| No vendor | 10.2 | 12.5 | 11.6 | 13.2 |

Source: GAO analysis of CMS data.

Note: Large vendors served more than 100 hospitals, medium vendors served 20 to 100 hospitals, and small vendors served fewer than 20 hospitals. Calculation of accuracy scores for the expanded-measure set was based on all the measures for which a hospital submitted data, which could range from the APU measures alone to a maximum of 17—the APU measures plus as many as 7 additional measures. CMS deems hospitals that achieve an accuracy score of 80 or better as having met its requirement to submit accurate data.

Rank ordering hospitals by the breadth of the confidence intervals around their accuracy scores, from the narrowest to the widest intervals, shows the large variation that we found across both quarters and measure sets. Hospitals with the narrowest confidence intervals, shown in table 9 as the 10th percentile, had a range of no more than 6 percentage points between the lower and upper limits of their confidence interval. That meant that their accuracy scores from one sample to the next were likely to vary by no more than plus or minus 3 percentage points from the accuracy score obtained in the sample drawn by CMS. By contrast, hospitals with the widest confidence intervals, shown in table 9 as the 90th percentile, exceeded 36 percentage points from the lower limit to the upper limit of their confidence interval. The accuracy scores for these hospitals would likely vary from one sample to the next by 18 percentage points or more, up or down, relative to the accuracy score derived from the CMS sample. For hospitals whose confidence interval included the 80 percent threshold, it was statistically uncertain whether a different sample of cases would have altered their result from passing the 80 percent threshold to failing, or vice versa.

**Table 9: Breadth of Confidence Intervals in Percentage Points Around the Hospital Baseline Accuracy Scores at Selected Percentiles, by Measure Set and Quarter**

| Hospital percentiles from narrowest to widest confidence intervals | APU-measure set | | Expanded-measure set | |
|---|---|---|---|---|
| | January-March 2004 discharges | April-June 2004 discharges | January-March 2004 discharges | April-June 2004 discharges |
| 10th percentile | 5.4 | 0.0 | 6.0 | 5.6 |
| 25th percentile | 8.1 | 7.3 | 9.3 | 8.2 |
| Median | 14.0 | 11.8 | 14.6 | 13.0 |
| 75th percentile | 24.2 | 21.5 | 23.6 | 21.3 |
| 90th percentile | 40.3 | 41.0 | 37.9 | 36.8 |

Source: GAO analysis of CMS data.

Note: Confidence interval based on a 95 percent significance level. Calculation of accuracy scores and confidence intervals for the expanded-measure set was based on all the measures for which a hospital submitted data, which could range from the APU measures alone to a maximum of 17—the APU measures plus as many as 7 additional measures.

One-third to one-fourth of hospitals had statistically uncertain results because their confidence interval extended both above and below the 80 percent threshold. Some of these hospitals had accuracy scores of 80 or above and some had scores of less than 80. Table 10 separates these hospitals into (1) those that had accuracy scores equal to 80 or above and were statistically uncertain and (2) those that had accuracy scores below 80 and were statistically uncertain. The table shows that most of the statistical uncertainty involved hospitals that passed CMS's accuracy threshold, but if a different sample of cases had been reabstracted by CDAC, there was a substantial possibility that they would not have passed.

**Table 10: For Hospitals with Confidence Intervals That Included the 80 Percent Threshold, Percentage of Total Hospitals with an Actual Baseline Accuracy Score That Either Met or Failed to Meet the Threshold, by Measure Set and Quarter**

| | APU-measure set | | Expanded-measure set | |
|---|---|---|---|---|
| | January-March 2004 discharges | April-June 2004 discharges | January-March 2004 discharges | April-June 2004 discharges |
| Percentage of hospitals whose actual accuracy score equals 80 or better | 23.9 | 19.2 | 28.0 | 24.0 |
| Percentage of hospitals whose actual accuracy score equals less than 80 | 8.3 | 7.0 | 11.3 | 8.7 |
| **Total** | **32.2** | **26.3** | **39.2** | **32.7** |

Source: GAO analysis of CMS data.

Note: Confidence interval based on a 95 percent significance level. Calculation of accuracy scores for the expanded-measure set was based on all the measures for which a hospital submitted data, which could range from the APU measures alone to a maximum of 17—the APU measures plus as many as 7 additional measures. CMS deems hospitals that achieve an accuracy score of 80 or better as having met its requirement to submit accurate data.

DEPARTMENT OF HEALTH & HUMAN SERVICES                    Centers for Medicare & Medicaid Services

*Administrator*
Washington, DC 20201

**DATE:**          DEC  - 9 2005

**TO:**            Cynthia A. Bascetta
                   Director, Health Care
                   Government Accountability Office

**FROM:**          Mark B. McClellan, M.D., Ph.D.
                   Administrator
                   Centers for Medicare & Medicaid Services

**SUBJECT:**       Government Accountability Office's (GAO) Draft Report: *HOSPITAL
                   QUALITY DATA: CMS Needs More Rigorous Methods to Ensure
                   Reliability of Publicly Released Data* (GAO-06-54)

Thank you for the opportunity to review and comment on the above-referenced subject
report. The Centers for Medicare & Medicaid Services (CMS) welcomes GAO's finding
of a high overall baseline accuracy rate when it examined CMS' assessment of the
hospital quality data submitted for the Annual Payment Update (APU) program. With
respect to GAO's finding that CMS has established no ongoing process to check data
completeness, we do have a process in place. CMS checked data completeness annually
during the 2 years of the APU program by assessing counts of hospital submitted data
relative to Medicare claims submissions for all hospitals included in the program.

The CMS instituted voluntary hospital reporting of quality data in 2003, for the first time,
as part of the Quality Improvement Organization (QIO) 7th Statement of Work. CMS
required the QIOs to assist all hospitals nationwide with voluntary reporting of a set of
quality measures to a clinical warehouse, and to help hospitals improve performance on
these measures. The quality measures covered four clinical topics; Acute Myocardial
Infarction, Heart Failure, Pneumonia, and Surgical Infection Prevention.

Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act
of 2003 dramatically changed the environment for hospital reporting. Per this provision,
Prospective Payment System (PPS) hospitals not submitting a set of 10 quality measures
receive a reduced APU by 0.4 percent for Fiscal Years 2005, 2006, and 2007.

The CMS successfully implemented and administered the APU program during 2004.
The number of hospitals submitting data to the clinical warehouse increased dramatically,
with nearly 99 percent of eligible hospitals participating. As a result of the APU, CMS
modified the reporting program by adding a validation component under which CMS
contractors assessed the accuracy of hospital chart abstraction for hospitals submitting
data. Additionally, CMS assessed relative volume of hospital reporting by comparing

Page 2 - Cynthia A. Bascetta

reporting volume by hospitals to their Medicare claims submissions. As a result of this successful implementation, public reporting of hospital quality data for about 3,600 hospitals on the Department of Health and Human Services Compare Web site was launched in 2004. Over 99 percent of PPS hospitals received the full APU payment in the 2004 determination.

In 2005, CMS successfully strengthened the accuracy criteria for the APU program. CMS expanded the APU criteria during 2005 to require hospitals with at least 6 eligible patients per quarter in the covered topics to submit accurate data. Through contractors, the CMS assessed the accuracy of submitted data, and determined that approximately 99 percent of submitting hospitals achieved an 80 percent upper bound of confidence interval accuracy threshold. The CMS used only 2 quarters of data ($3^{rd}$ and $4^{th}$ quarter 2004 discharges) because previous quarters' measures definitions were not completely aligned with the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) measures. Prior to this reporting period, these measure definitional differences slightly impacted some JCAHO hospitals' accuracy results.

Additionally in 2005, the CMS expanded its review of the completeness of data submissions by comparing the volume of submissions to claims data by quarter and by topic. CMS found that discrepancies between the volume of claims and data submissions in about 20 percent of hospitals. However, 96 percent of PPS hospitals submitted data for topics for which they had eligible cases for each quarter, and therefore received the full update in the 2005 determination.

The CMS appreciates the thoughtful analysis and recommendations in the GAO report. The Agency believes that its methods to evaluate accuracy of submissions are sound, and agree that the quality and completeness of the data must be improved. This can best be accomplished through quarterly reports to hospitals that promote improvement in data accuracy. The Agency is also considering various other steps, as indicated in the detailed comments to the recommendations.

Attached are the detailed comments to the GAO's recommendations.

**Centers for Medicare & Medicaid Services' (CMS) Comments to the Government Accountability Office's (GAO) Draft Report:** *HOSPITAL QUALITY DATA: CMS Needs More Rigorous Methods to Ensure Reliability of Publicly Released Data* **(GAO-06-54)**

In order for CMS to help ensure the reliability of the quality data it uses to produce information on hospital performance, GAO wrote three recommendations to CMS. Our responses follow each recommendation below:

**GAO Recommendation**
Focusing on the subset of hospitals for which it is statistically uncertain if they met CMS accuracy threshold in one or more previous quarters, increase the number of patient records reabstracted by the clinical data abstraction center so that the proportion of hospitals with statistically uncertain results is reduced.

**CMS Response**
The CMS believes that our methods to evaluate accuracy of submissions are sound. Our quarterly reabstraction sample of hospital submitted data found that hospital submitted data were generally accurate, as evident by approximately 90 percent of submitting hospitals achieving the 80 percent accuracy threshold. We agree that the quarterly accuracy estimates using 5 sample charts can have considerable sampling error, and is highly dependent on the clustering of errors within individual charts. As GAO cited in the report, confidence intervals generally range from 10 percent to 14 percent, but can be as high as 35 percent to 40 percent when errors are clustered in a single chart.

Measure definition differences with the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) prevented CMS from using accuracy results prior to 3rd quarter 2004 for the 2005 Annual Payment Update (APU) determination, since slight definitional differences might differentially impact accuracy rates. These definitional differences were resolved with the July 2005 and January 2006 alignment modifications, where all hospitals submit data for the same measures.

The quarterly accuracy estimates are primarily designed to provide hospitals and their vendors periodic feedback about the relative accuracy of their abstraction processing for quality improvement. We recommend that hospitals and vendors analyze accuracy results from multiple quarters to assess their abstraction processes. We will continue to educate hospitals and vendors to analyze all quarters' reabstraction accuracy results in order to provide a more reliable snapshot of their abstraction accuracy over time.

Timing issues prevent CMS from implementing a targeted reabstraction subsampling of hospitals falling into the statistically uncertain outcome range. Identifying these hospitals requires completion of the initial 90 day reabstraction and appeals process. Selecting additional records after initial validation results are available would require an additional 2 to 3 months to select the sample, request the additional charts, provide sufficient time for hospitals to identify and send the requested charts, reabstract the data elements, and provide additional time for appeals processing.

Page 2 - Attachment

For the fiscal year (FY) 2007 APU determination, CMS will improve the stability of accuracy estimates by using all 4 quarters' validation estimates to provide a more stable estimate of their abstraction accuracy. As pointed out in the report, combining several quarters' accuracy estimates will provide a more stable estimate with lessened likelihood of clustering effects impacting sampling variability. This combining of multiple quarters' data will dramatically decrease the number of hospitals with statistically uncertain results for the FY 2007 annual payment update determination.

**GAO Recommendation**
Require hospitals to certify that they took steps to ensure that they submitted data on all eligible patients, or a representative sample thereof.

**CMS Response**
As noted in the report, CMS assessed the completeness of hospital submitted data for Medicare beneficiaries in the FY 2005 and 2006 APU determinations. CMS found during the FY 2006 APU determination that about 20 percent of the prospective payment system hospitals submitted fewer patient records than claims submissions. CMS will develop stronger methods of assessing the completeness of hospital data submissions. We will implement two requirements:

1. Require that hospitals formally attest to the completeness of their quarterly submission of quality data; and
2. Require that hospitals submit an aggregate count of all eligible Medicare and non-Medicare patients.

**GAO Recommendation**
Assess the level of incomplete data submitted by hospitals for the APU program to determine the magnitude of underreporting, if any, in order to refine how completeness assessments may be done in future reporting efforts.

**CMS Response**
As noted in the GAO report above, CMS assessed completeness by using Medicare claims data to compare to hospital submissions.

As a result of implementing the two requirements described above, we will have data that enable us to assess completeness of hospital submissions for Medicare patients by comparing the number of submissions to the quality improvement organization clinical warehouse with both claims submissions and hospital-submitted eligible patient counts. For non-Medicare patients, the number of warehouse submissions will be compared with hospital-submitted patient counts. Hospitals will be asked to explain discrepancies. Based on this information, CMS will be able to consider whether it is necessary to take additional steps to assess and improve the completeness of submissions.

Page 3 - Attachment

The CMS believes that our methods for assessing abstraction accuracy and completeness of Medicare beneficiary submissions are basically sound. Again, we appreciate the insightful analysis and recommendations that GAO has provided CMS in this report. Among the steps that CMS will implement in improving our hospital reporting program are the following:

- Combine several quarters of accuracy estimates to provide a more stable estimate of accuracy of chart abstraction;
- Require that hospitals formally attest to the completeness of their quarterly submission of quality data;
- Require that hospitals submit an aggregate count of all eligible Medicare and non-Medicare patients;
- Analyze completeness of the hospital patient data submission by comparing submission counts with counts of claims submission and eligible individuals who are Medicare patients, and require hospitals to explain discrepancies; and
- Continue to provide quarterly feedback to hospitals about submission accuracy and completeness, and require them to explain discrepancies among counts.

# Appendix V: GAO Contact and Staff Acknowledgments

## GAO Contact

Cynthia A. Bascetta (202) 512-7101 or BascettaC@gao.gov

## Acknowledgments

In addition to the contact named above, Linda T. Kohn, Assistant Director; Ba Lin; Nkeruka Okonmah; Eric A. Peterson; Roseanne Price; and Jessica C. Smith made key contributions to this report.

| | |
|---|---|
| **GAO's Mission** | The Government Accountability Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability. |
| **Obtaining Copies of GAO Reports and Testimony** | The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday, GAO posts newly released reports, testimony, and correspondence on its Web site. To have GAO e-mail you a list of newly posted products every afternoon, go to www.gao.gov and select "Subscribe to Updates." |
| **Order by Mail or Phone** | The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:<br><br>U.S. Government Accountability Office<br>441 G Street NW, Room LM<br>Washington, D.C. 20548<br><br>To order by Phone:  Voice:  (202) 512-6000<br>TDD:  (202) 512-2537<br>Fax:  (202) 512-6061 |
| **To Report Fraud, Waste, and Abuse in Federal Programs** | Contact:<br><br>Web site: www.gao.gov/fraudnet/fraudnet.htm<br>E-mail: fraudnet@gao.gov<br>Automated answering system: (800) 424-5454 or (202) 512-7470 |
| **Congressional Relations** | Gloria Jarmon, Managing Director, JarmonG@gao.gov (202) 512-4400<br>U.S. Government Accountability Office, 441 G Street NW, Room 7125<br>Washington, D.C. 20548 |
| **Public Affairs** | Paul Anderson, Managing Director, AndersonP1@gao.gov (202) 512-4800<br>U.S. Government Accountability Office, 441 G Street NW, Room 7149<br>Washington, D.C. 20548 |